

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2004, 10:03:59 ; Search time 28 Seconds
(without alignments)
559.108 Million cell updates/sec

Title: US-10-081-644-2

Perfect score: 1920

Sequence: 1 MSVPTTQKAVIEGDKAVK.....ALTEGIEKGNKNVYARL 370

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PTCUS COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	266.5	13.9	363	4	US-09-215-694-5
2	202.5	10.5	340	4	US-09-107-532A-7298
3	173.5	9.0	350	4	US-09-328-352-8160
4	158.5	8.3	396	3	US-08-860-656B-2
5	158.5	8.3	396	3	US-08-860-656B-3
6	153.5	8.0	336	1	US-08-713-254-2
7	153.5	8.0	336	3	US-08-855-767-2
8	151.5	7.9	340	4	US-09-134-001C-3258
9	149.5	7.8	322	4	US-09-154-750A-81
10	141	7.3	349	4	US-09-107-532A-5838
11	139.5	7.3	344	4	US-09-252-991A-16756
12	134	7.0	401	4	US-09-252-991A-19274
13	133	6.9	436	4	US-09-410-551B-72
14	131.5	6.8	396	3	US-08-860-656B-4
15	130.5	6.8	370	4	US-09-328-352-5285
16	126	6.6	2532	4	US-09-215-694-10
17	125.5	6.5	336	4	US-09-252-991A-11476
18	124.5	6.5	371	4	US-08-501-115-14
19	122.5	6.4	436	4	US-08-976-063E-20
20	121.5	6.3	412	4	US-09-252-991A-22667
21	121	6.3	345	2	US-07-857-224B-94
22	118	6.1	344	2	US-07-857-224B-93
23	115.5	6.0	2152	3	US-09-036-987A-3
24	115.5	6.0	2152	3	US-09-370-700-3
25	115.5	6.0	2152	4	US-09-603-207-3
26	114	5.9	3724	2	US-08-804-227C-10
27	114	5.9	3724	2	US-08-804-198-4

ALIGNMENTS

RESULT 1
US-09-215-694-5
; Sequence 5, Application US/09215694B
; Patent No. 6391583
; GENERAL INFORMATION:

Sequence 5638, Ap
Sequence 2, Appli
Sequence 10, Appli
Sequence 29389, A
Sequence 92, Appli
Sequence 95, Appli
Sequence 32, Appli
Sequence 4, Appli
Sequence 8, Appli
Sequence 3, Appli
Sequence 7738, Ap
Sequence 3374, Ap
Sequence 216, App
Sequence 216, App
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 14, Appli
Sequence 28, Appli
Sequence 96, Appli
Sequence 24660, A
Sequence 4988, Ap
Sequence 2, Appli
Sequence 32383, A
Sequence 1085, Ap
Sequence 2, Appli
Sequence 2, Appli
Sequence 7360, Ap
Sequence 34, Appli
Sequence 89, Appli
Sequence 31262, A
Sequence 2, Appli
Sequence 33, Appli
Sequence 2, Appli
Sequence 3, Appli
Sequence 6054, Ap
Sequence 5, Appli
Sequence 5, Appli
Sequence 5, Appli
Sequence 5, Appli
Sequence 5, Appli
Sequence 7, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 5343, Ap
Sequence 12, Appli
Sequence 4, Appli
Sequence 4056, Ap
Sequence 1, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 5503, Ap
Sequence 7473, Ap
Sequence 6, Appli
Sequence 124, App
Sequence 25683, A
Sequence 4, Appli

Query Match 9.0%; Score 173.5; DB 4; Length 350;
Best Local Similarity 29.4%; Pred. No. 3.4e-09;
Matches 77; Conservative 34; Mismatches 106; Indels 45; Gaps 12;

Qy 1 MSVPTTKAVIIEGDKAVKTDVSVPELKEGTALVYVEAVAGNPTDVKHIAIKIGPESG- 59
Db 15 MKAVAYKAGFPITLPEALVDIELTTPVAKGHDLVVRQAVSNVPTDK-IRKNVSARQSG 73
Qy 60 --ILGCDIAGTVKLGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAYARVYPPLYKS 117
Db 74 WKVLGWDVAGTVKIGV--TOFKIGV--VWYAGALN---RQGSNELQLV----- 119
Qy 118 NLTHSTADE--ISBGPVKNFES--AASLPVSLTTAGVSLCHLGSKMWHPTSTPQHTHPL 173
Db 120 -----DERIVGHKP-KTLEATEAALPLTALTAWEMLPDLQV-----PKTAPANTTI 166
Qy 174 LINGGATAVGQQLIQVAKHINAYTKIVTVASKKHEKLLKSYGADDVDFYHD--AGVIEQI 231
Db 167 LVIGGAGGVGSITTLQLLKQLTNLTIIATASRPETKEWVEQLGADYVLDHRQPLAAQIKOL 226
Qy 232 KSKYP-----NLQHVIVDAV 245
Db 227 GLSAPLVFSTTDDQLSLIV 248

RESULT 4
US-08-860-656B-2
; Sequence 2, Application US/08860656B
; Patent No. 6297028
; GENERAL INFORMATION:
; APPLICANT: Taniguchi, Tadatsugu
; APPLICANT: Shibuya, Hiroshi
; APPLICANT: Barsoumian, Edward L.
; TITLE OF INVENTION: IL-2R-Associated Polypeptides and DNA Molecules Coding Therefor
; FILE REFERENCE: 0652.1640000
; CURRENT APPLICATION NUMBER: US/08/860, 656B
; PRIOR FILING DATE: 1998-01-15
; PRIOR FILING DATE: 1995-12-23
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 396
; TYPE: PRT
; ORGANISM: Human p43
US-08-860-656B-2

Query Match 8.3%; Score 158.5; DB 3; Length 396;
Best Local Similarity 27.8%; Pred. No. 1.4e-07;
Matches 71; Conservative 39; Mismatches 96; Indels 49; Gaps 12;

Qy 34 LVKVEAVAGNPTDVK-----HIAYKIGPESGI-LGCDIAGTVVVKLGN 75
Db 75 IVKHAASVNPIDVNRSGYGATALNMKRDPLHVKIK-GEFFPLTLGRDVGVMVECGLD 133
Qy 76 ASTDLKVGDTGFGFVHGASQTDPKNGAFAYARVYPPLYKSNLTHSTADEISBGPVK-N 134
Db 134 VKY-FKPGDEVM-----AAVPPWKQGLTSEFVV-----SGNEVSHKPKSLT 174
Qy 135 FESAASLPVSLTTAGVSLCHLGSKMWHPTSTPQHTHPLINGGATAVGQQLIQVAKHIN 194
Db 175 HTQAASLPYVALTAW-SAINKVGGLNDKNCCTGKR-----VLILGASGGVGTFAIQVMKAWD 229
Qy 195 AYTKIVTVASKKHEKLLKSYGADDVDFYHDAGVIEQIKSKYPNLOHVIVDAVG--SEDSIP 252
Db 230 AH--VTAVCSQDASELVRKLGADDVIDYKSGSVEEQLKSLKP-PDFILDNVGGSTETWAP 286
Qy 253 EAYKVTADSLPATLL 267
Db 287 DFLKWSGATVTVTLV 301

RESULT 5
US-08-860-656B-3
; Sequence 3, Application US/08860656B
; Patent No. 6297028
; GENERAL INFORMATION:
; APPLICANT: Taniguchi, Tadatsugu
; APPLICANT: Shibuya, Hiroshi
; APPLICANT: Barsoumian, Edward L.
; TITLE OF INVENTION: IL-2R-Associated Polypeptides and DNA Molecules Coding Therefor
; FILE REFERENCE: 0652.1640000
; CURRENT APPLICATION NUMBER: US/08/860, 656B
; CURRENT FILING DATE: 1998-01-15
; PRIOR FILING DATE: 1995-12-23
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 3
; LENGTH: 396
; TYPE: PRT
; ORGANISM: Human p43
US-08-860-656B-3

Query Match 8.3%; Score 158.5; DB 3; Length 396;
Best Local Similarity 27.8%; Pred. No. 1.4e-07;
Matches 71; Conservative 39; Mismatches 96; Indels 49; Gaps 12;

Qy 34 LVKVEAVAGNPTDVK-----HIAYKIGPESGI-LGCDIAGTVVVKLGN 75
Db 75 IVKHAASVNPIDVNRSGYGATALNMKRDPLHVKIK-GEFFPLTLGRDVGVMVECGLD 133
Qy 76 ASTDLKVGDTGFGFVHGASQTDPKNGAFAYARVYPPLYKSNLTHSTADEISBGPVK-N 134
Db 134 VKY-FKPGDEVM-----AAVPPWKQGLTSEFVV-----SGNEVSHKPKSLT 174
Qy 135 FESAASLPVSLTTAGVSLCHLGSKMWHPTSTPQHTHPLINGGATAVGQQLIQVAKHIN 194
Db 175 HTQAASLPYVALTAW-SAINKVGGLNDKNCCTGKR-----VLILGASGGVGTFAIQVMKAWD 229
Qy 195 AYTKIVTVASKKHEKLLKSYGADDVDFYHDAGVIEQIKSKYPNLOHVIVDAVG--SEDSIP 252
Db 230 AH--VTAVCSQDASELVRKLGADDVIDYKSGSVEEQLKSLKP-PDFILDNVGGSTETWAP 286
Qy 253 EAYKVTADSLPATLL 267
Db 287 DFLKWSGATVTVTLV 301

RESULT 6
US-08-713-254-2
; Sequence 2, Application US/08713254
; Patent No. 5763236
; GENERAL INFORMATION:
; APPLICANT: KOJIMA, TOMOKO
; APPLICANT: YAMAMOTO, HIROAKI
; APPLICANT: KAWADA, NAOKI
; APPLICANT: MATSUYAMA, AKINOBU
; TITLE OF INVENTION: NOVEL ENZYME, A METHOD TO PREPARE SAID
; TITLE OF INVENTION: ENZYME, A DNA SEGMENT ENCODING SAID ENZYME, A TRANSFORMANT
; TITLE OF INVENTION: CONTAINING SAID DNA SEGMENT AND A METHOD OF PREPARING
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/713,254
FILING DATE: 12-SEP-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/311,328
FILING DATE: 23-SEP-1994
APPLICATION NUMBER: JP 5-261649
FILING DATE: 24-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 5-337191
FILING DATE: 28-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-181308
FILING DATE: 02-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Oblon, No. 5763236man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 6423-001-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 336 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-713-254-2

Query Match 8.0%; Score 153.5; DB 1; Length 336;
Best Local Similarity 24.8%; Pred. No. 3.5e-07;
Matches 101; Conservative 50; Mismatches 125; Indels 131; Gaps 22;
Qy 1 MSVPTTQKAVIIEGDKAV-VKTDVSVPELKEGTALVKVEAVAGNPTDWHKAIYK---IGP 56
Db 1 MSIPSSQYGFVFNKQSGNLNRNDLPVHKPKAGQLLLKVDVAGLCHSD-LHVIYEGDLCG- 58
Qy 57 EGSILGCDIAGTVVKLGPNASTDLKVGDT-----GFGFVHGASQTDPKN----- 100
Db 59 DNYVMGHEIAGTAAVAGDDV-INYKVGDRVACVGNCGCGCKYCRGAIDNVCKNAFGDWF 117
Qy 101 -----GAFAEYARVYPPFLFYKSNLTHSTADEISEGPKVNFESAASLPVSLTTAGVSLCHH 155
Db 118 GLGYDGGYQQYLLVTRP-----RNLSR-IPDNVSAD-----VAASTDAVLTP-----YH 161
Qy 156 LGSKMEWHPSTPQTHPLIIGWGATAVGQOLLIQVAKHINAYTKIVTVASKKHE--KLLKS 213
Db 162 AIKMAQVSP-----SNILLI--GAGGLGNAIQVAK--AFGAKVTVLDDKKEARDOAKK 212
Qy 214 YGADVDYHDAGVIEQIKSKYPNLQHVDAVSEDSIPEAYKVTADSLPATLLEVVPM 273
Db 213 LGADAVY----- 219
Qy 274 IESIPETIRKDNVKIDITLLYRASQGEILLGATFRFPASPEVHE--ATVKFVKFINPHL-- 329
Db 220 -ETLPESISPGS-----FSACDFVSVQAT-FDVCQYKVPKGVMPVGLGAPNLSF 269
Qy 330 NNGDIHMNIKVF-----SNGLDDVPALTEGIEKGNKNKVKVVARL 370
Db 270 NLGDLALREIRILGSGFWGTNDLDDVLKL---VSEGVKVPVRSACL 313

RESULT 7
US-08-855-767-2
Sequence 2, Application US/08855767
Patent No. 6255092
GENERAL INFORMATION:
APPLICANT: KOJIMA, TOMOKO
APPLICANT: YAMAMOTO, HIROAKI
APPLICANT: KAWADA, NAOKI

APPLICANT: MATSUYAMA, AKINOBU
TITLE OF INVENTION: NOVEL ENZYME, A METHOD TO PREPARE SAID
TITLE OF INVENTION: ENZYME, A DNA SEGMENT ENCODING SAID ENZYME, A TRANSFORMANT
TITLE OF INVENTION: CONTAINING SAID DNA SEGMENT AND A METHOD OF PREPARING
TITLE OF INVENTION: OPTICALLY ACTIVE ALCOHOL USING SAID ENZYME
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESS: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/855,767
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/311,328
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 5-337191
FILING DATE: 28-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-181308
FILING DATE: 02-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Oblon, No. 6255092man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 6423-001-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 336 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-855-767-2

Query Match 8.0%; Score 153.5; DB 3; Length 336;
Best Local Similarity 24.8%; Pred. No. 3.5e-07;
Matches 101; Conservative 50; Mismatches 125; Indels 131; Gaps 22;
Qy 1 MSVPTTQKAVIIEGDKAV-VKTDVSVPELKEGTALVKVEAVAGNPTDWHKAIYK---IGP 56
Db 1 MSIPSSQYGFVFNKQSGNLNRNDLPVHKPKAGQLLLKVDVAGLCHSD-LHVIYEGDLCG- 58
Qy 57 EGSILGCDIAGTVVKLGPNASTDLKVGDT-----GFGFVHGASQTDPKN----- 100
Db 59 DNYVMGHEIAGTAAVAGDDV-INYKVGDRVACVGNCGCGCKYCRGAIDNVCKNAFGDWF 117
Qy 101 -----GAFAEYARVYPPFLFYKSNLTHSTADEISEGPKVNFESAASLPVSLTTAGVSLCHH 155
Db 118 GLGYDGGYQQYLLVTRP-----RNLSR-IPDNVSAD-----VAASTDAVLTP-----YH 161
Qy 156 LGSKMEWHPSTPQTHPLIIGWGATAVGQOLLIQVAKHINAYTKIVTVASKKHE--KLLKS 213
Db 162 AIKMAQVSP-----SNILLI--GAGGLGNAIQVAK--AFGAKVTVLDDKKEARDOAKK 212
Qy 214 YGADVDYHDAGVIEQIKSKYPNLQHVDAVSEDSIPEAYKVTADSLPATLLEVVPM 273
Db 213 LGADAVY----- 219
Qy 274 IESIPETIRKDNVKIDITLLYRASQGEILLGATFRFPASPEVHE--ATVKFVKFINPHL-- 329

```
Db 220 -ETLPESIPGS-----FSACFDVSVQAT-FDVCQKYVEPKGVIMPVGLGAENLSF 269
QY 330 NNGDIHMMIKVF-----SNGLDDVPALTEGKEGKNKVKVVARL 370
Db 270 NUGDLALREIRILGSPWGTNDLDDVLKL---VSEKVKPVVRSACL 313

RESULT 8
US-09-134-001C-3258
; Sequence 3258, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3258
; TYPE: PR
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3258

Query Match
Best Local Similarity 7.9%; Score 151.5; DB 4; Length 340;
Matches 67; Conservative 33; Mismatches 102; Indels 39; Gaps 11;

QY 6 TQKAVIIEGKAVKTD-----VSVPDELKEGTALVKVEAVAGNPTDKHIAKIGP 56
Db 4 TWKAI---GFKSSFQLDGECNCFEENFDIPHPGSHLLVKVQISVNPVDTKQRTMPVK 60
QY 57 EGSILGCDIAGTVVKLGNASTDLTKVGDGTGFGFVHGASQTDPKNGAFAYARVVPPLFYK 116
Db 61 APRVLGFDVGVIEKIQDQVSM-FQEGDVVF-----YSGSPNQNGSNEEYQLIEEYL 111
QY 117 SNLTHSTADEISSEGPVK-NFESAASLPVSLTTAGVSLCHHLGSKMEWHPSTPQHTHPLLI 175
Db 112 -----VAKAPTNLSEQAASLPGLTGLTAVETLFDVFGISKE--PS-ENKGSLLI 158
QY 176 WGGATAVGQQLIQVAKHINATKIYTVASKKHE-KLLKSYGADDVFDYHDAGVIEQIKSK 234
Db 159 INGAGGVGIATQIAKFG--LKVITTSAREDTIKWSVNNMGADVVLN-HKKDLISQOFKON 215
QY 235 Y 235
Db 216 H 216

RESULT 9
US-09-154-750A-81
; Sequence 81, Application US/09154750A
; Patent No. 6432640
; GENERAL INFORMATION:
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Polyak, Kornelia
; TITLE OF INVENTION: p53-Induced Apoptosis
; FILE REFERENCE: 1107.75357
; CURRENT APPLICATION NUMBER: US/09/154,750A
; CURRENT FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/059,153
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/079817
; PRIOR FILING DATE: 1998-03-30
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 81
; LENGTH: 322
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-154-750A-81

Query Match
Best Local Similarity 7.8%; Score 149.5; DB 4; Length 322;
Matches 64; Conservative 34; Mismatches 105; Indels 35; Gaps 9;

QY 14 GDKAVVKTDSVPELKEGTALVKVEAVAGNPTDKHIAKIGP---EGSILGCDIAGTVV 70
Db 11 GPNLYVKEVAKESPGESEVLLKVAASALNRADLMORQGOYDPPPGASNLGLEASGHA 70
QY 71 KLGPNASTDLTKVGDGTGFGFVHGASQTDPKNGAFAYARVVPPLFYKSNLTHSTADEISEG 130
Db 71 ELGPGCGQHWKIGDNTAMALLPGGQ-----AQYTVPEGLM-----PIPEG 112
QY 131 PVKNFESAASLPVSLTTAGVSLCHHLGSKMEWHPSTPQHTHPLLIWGGATAVGQQLIQVA 190
Db 113 --LTLTQAAAIPAEMLTA-FQLLHLVGN-----VQAGDYVLIHAGLSGVGTAAIQLT 161
QY 191 KHINATKIYTVASKKHEKLLKSYGADDVFDYHDAGVIE-QIK-SKYPNLOHVIDAVG 246
Db 162 RMAGA-IPLVTAGSQKQLQMAEKLGAAAGFNKYKEDFSEATLKTGAGVNLIDLCIG 218

RESULT 10
US-09-107-532A-5838
; Sequence 5838, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; CORRESPONDENCE ADDRESS: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5838:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 349 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
```

NAME/KEY: misc.feature
LOCATION: (B) LOCATION 1...349
SEQUENCE DESCRIPTION: SEQ ID NO: 5838:
US-09-107-532A-5838

Query Match 7.3%; Score 141; DB 4; Length 349;
Best Local Similarity 22.5%; Pred. No. 6.9e-06;
Matches 89; Conservative 52; Mismatches 148; Indels 106; Gaps 17;

QY 1 MSVPTTKAVILBGRKAVKTDVSVPELKEGTALVKVEAVAGNPTDKHIAIKIGEGS- 59
DB 11 MKVVGFEGLPIDDPKSFIDEKQIPVPSRDLLVKVAVSNVPTK-LRQDNGIRNAL 69
QY 60 -ILGCDIAGTVKLGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAYEARVYPLPVKSN 118
DB 70 RIILGFGVGVKVAAGVEQV-KFSVGDRVP-----YAGTTTRAGSNQEQYQLV----- 114
QY 119 LTHSTADE--ISRGPKVNF--ESAASLPVSLTTAGVSLCHLHSGKMEWHPSTPQHT- 170
DB 115 -----DERIVALAP-KNLSDEEAALPLTSLTAYELLPEKFG-----LTPENANRG 160
QY 171 HPLLINGGATAVGQOLIOVAKHINAYTKIVTVASKKHEKLLKSYGADDVFDYHD- 224
DB 161 KKLIVNGSGGVGSLNQLAHW--AGLEVYATASPKNFELWKTGYDYPIDYHDLKSGF 218
QY 225 -----AGVIEQIKSKYPNQLQHVDAVGSSEDSIPEAYKVTADSLPATLLEVVPMTI 274
DB 219 RELGIDQVEVAVLFDITRFDQIKHLRPFQVHG-----AKTDYNYETIQTGEALAHIA 255
QY 275 ESIPERI-----RDNVKKIDITILYRAGQOELLGATRRPPASPEYH---EATVKFVKPFIN 326
DB 256 VGIEKPLDIDGKWNQISPDWEYMF-----AKTDYNYETIQTGEALAHIA 300
QY 327 PHLNGGDIHWNKIVPSNGLD--DVPALTEGIKEG 359
DB 301 ALANEGIIIRSTLKVSYSDGNKAKNLQKATKDVETG 335

RESULT 11

US-09-252-991A-16756

Sequence 16756, Application US/09252991A
Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 16756

LENGTH: 344

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-16756

Query Match 7.3%; Score 139.5; DB 4; Length 344;

Best Local Similarity 26.9%; Pred. No. 9.6e-06;

Matches 59; Conservative 35; Mismatches 90; Indels 35; Gaps 10;

QY 22 DVSVPKLEK--TALVKVEAVAGNPTDKHIAIKIGP---GSLGCDIAGTVKLGPN 76
DB 27 DLSLPEPTGPRDLLVEVAISVNPVDTK-VROQAQPEAGQAKVGLWDAAGVVRVAGSEV 85

QY 77 STLKVGDTGFGFVHGASQTDPKNGAFAYEARVYPLPVKSNLTHSTADEISGPVK--- 133
DB 86 SL-FRPGDRVW---YAGDITRP--GNSSELHRV-----DERIAGHLPKSL 124

QY 134 NFESAASLPVSLTTAGVSLCHLHSGKMEWHPSTPQHTPLLIWGATAVGQOLIOVAKHI 193

DB 125 DFAQAALPLTTITA-----WELLFERLQIAEGKADQGSLLVVGAGGVGSILVQLARQL 180
QY 194 NAYTKIVTVASKKHEKLLKSYGADDVFDYHDAGVIEQIK 232
DB 181 TGLNVIGTASRAETQAWVRDLGAHVID-HGKPLABELK 218

RESULT 12

US-09-252-991A-19274

Sequence 19274, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 19274

LENGTH: 401

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-19274

Query Match 7.0%; Score 134; DB 4; Length 401;

Best Local Similarity 21.3%; Pred. No. 4.5e-05;

Matches 73; Conservative 54; Mismatches 150; Indels 66; Gaps 13;

QY 14 GDKAVKTDVSVPELKEGTALVKVEAVAGNPTDKHIAIKIG-----PEGS--ILGCDIAG 67
DB 84 GPEVLVATSRPLTPGPREVLVEVRAAGVNGPD---VLQRKGVYDPPPGASDIPGLEIAG 140

QY 68 TVVKLGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAYEARVYPLPVKSNLTHSTADE- 126
DB 141 MYRANGSEVSR-FAVGEAVMALIPG-----GGYAQFA-----VADER 176

QY 127 ----ISEGPKVKNFESAASLPVSLTTAGVSLCHLHSGKMEWHPSTPQHTPLLIWGATAV 182
DB 177 TTLHLPDG--LCWEERAALPEFTMTVWNLFORGGFKA-----GETLLVHGASGI 225

QY 183 GQOLIOVAKHINAYTKIVTVASKKHEKLLKSYGADDVFDYHDAGVIEQI--KSKYPNLOH 240
DB 226 GTAATMLGKAFGAAKIFTTISSEAQREASRLGADLAINYTEQDFVEEVLRTGREGVDV 285

QY 241 VIDAVGSEDSIPEAYKVTADSLPATLLEVVPMTIESIPEIRKDNVKIDITILYRASGOE 300
DB 286 IVDIVAG-DYVTRNYQAAAMNGRIVOIGVI-----KGKAAEVDLFFPMLSKRL 331

QY 301 ILLGAT-REPPASPEYHEATVKVFPINPHLNNGDIHNMNIKVF 342
DB 332 VHLGSTLSRSRSHDEKGAITIAELERQVWPHVRAGAVRPQVFRTP 374

RESULT 13

US-09-410-551B-72

Sequence 72, Application US/09410551B

Patent No. 6503737

GENERAL INFORMATION:

APPLICANT: KOSAN BIOSCIENCES, INC.

APPLICANT: REEVES, CHRISTOPHER

APPLICANT: CHU, DANIEL

APPLICANT: KHOSLA, CHAITAN

APPLICANT: SANTI, DANIEL

APPLICANT: WU, KAI

TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA

TITLE OF INVENTION: CONSTRUCTS THEREFOR

FILE REFERENCE: 30062-20026.00

; SEQ ID NO 10
; LENGTH: 2532
; TYPE: PRT
; ORGANISM: Aspergillus terreus
US-09-215-694-10

Query Match 6.6%; Score 126; DB 4; Length 2532;
Best Local Similarity 23.4%; Pred. No. 0.0056;
Matches 56; Conservative 39; Mismatches 92; Indels 52; Gaps 10;

Qy 18 VVKTDVPELKEGTALVK--VEAVAGN--PTDW-----KHIAKIGP-EGSI 60
Db LRLRDIQTPGLDLSLHFTKRNVDTEPKLDDWVEIEPRAFGNLFRIIMVAMQLESNV 1875
Qy 61 LGCDIAGTVVKLGNAST--DLKVGDTGFGFVHGASQTDPKNGAFAYARVPPLEPKS 117
Db 1876 MGFCAGVVTLSLSTARTIAPGLAVGDRVCALMNG-----HWAS 1914
Qy 118 NLHSTADEISEGPKVKNFESASLPVSLTTAGVSLCHHLGSKMEWHPSTPOHTHPLIWG 177
Db 1915 RVTTSRNVVRIPETLSFPHPAASIPLAFTAYISL--YTARI-----LPGET--VLIHA 1965
Qy 178 GATVAGQOLIQAQKHINAYTKIVTASKEKLLKSY--GADVDVYHDAGVIEQIKSK 234
Db 1966 GAGVGQAAILAQLTGAEVFTTAGSETKRNLLIDKFLDPDHVFSRDSFVDGIKTR 2024

RESULT 17
US-09-252-991A-31476
; Sequence 31476, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31476
; LENGTH: 336
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31476

Query Match 6.5%; Score 125.5; DB 4; Length 336;
Best Local Similarity 24.4%; Pred. No. 0.00025;
Matches 76; Conservative 41; Mismatches 99; Indels 95; Gaps 17;

Qy 31 GTALVKVEAVAGNPTDKHIAKIG-----PEGSILGCDIAGTVVKLGNASTDLKVGDTG 86
Db 41 GEVWLEQAAGVNFLD---LQORSQAVPIPLPSGLGLEGAGVVAALGPGVS--GLAPCD-- 94
Qy 87 FGFVHGASQTDPKNGAFAYARVPPLEPKSNTLHSTADEISEGP--VKNFESASLPVSL 145
Db 95 ----RVAYATGPL--GAYAS--ARLYP-----AERLLKLPDTLAFEDAAALFKG 136
Qy 146 TTAGVSLCHHLGSKMEWHPSTPOHTP-----LLIWWGATVAGQOLIQAQKHINAYTKI 199
Db 137 ITA-----HYL-----LYATIPVPGFTGRTLLYGAAGVAGQQLMAAWARHLGAW--V 179
Qy 200 VTVASKKHE--KLLKSYGADVDVYHDAGVIEQI-----KS 233
Db 180 IGUVSKAESVERAAGCDEVLFVDAASLAQAQVAELTAGRKVDVVDYDPIGRATFEASLNS 239
Qy 234 KYPNLOHVIDAVGESDIPKAYKVTADSLPATLLEVVPMTI--ESPEERKONVKIDIT 291
Db 240 LRP--RGLLVFGATGVPVPAVEATLNAGSLFLTRPSLAHAHTANPEE-----286

Query Match 6.5%; Score 124.5; DB 4; Length 371;
Best Local Similarity 22.0%; Pred. No. 0.00036;
Matches 74; Conservative 54; Mismatches 149; Indels 59; Gaps 15;

Qy 46 DKHIAKIGPEGSILGCDIAGTVVKLGNASTDLKVGDTGFGFVHGASQTDPKNGAF 103
Db 65 DKNAKYPMP-----GHEIAGEVTEVGRNV-TKFKAGDRVGVGCMVNSQSCSCDKGF 118
Qy 104 AEYARVYPLFYKSN-----LTH---STADEISEGPKVKNFESASLP--VSLTTAGVS 151
Db 119 ENHC---PGMIFTYNSVDRDGTTHGGYSSVWVHERFVVRFPDAMPDLKGPALLCAGIT 175
Qy 152 LCHHLGSKMEWHPSTPOHTHPLIWWGATVAGQOLIQAQKHINAYTKIVTASKEKLL 211
Db 176 VY-----SPMKYTHGLNAPGMHLGVL--GLGGLGHVAVKFGKAFKMTVVISSSPGKQEAL 229
Qy 212 KSYGADVDVYHDAGVIEQIKSKYPNLOHVIDAVGESDIPSA-----YKVTADSL 262
Db 230 ERIGADAFVVSQDA---DEMKATMTMDGIINTVSA--NVPWAPLEGLKPKNGKIMVGL 284
Qy 263 PATLLEVPWMTIESPEERKONVKIDITLLYRAG-----OEILIGATRFPPASPEYHEA 317
Db 285 PEKPIEVPPFAL-----VARNKTLAGSCIGGMRDTQEMLDLAAKHGVTADIEVI 333
Qy 318 TVKFKVFINPHLNGDIHMHNIKVFNSGLDDVPALT 353
Db 334 GAQYVNTAMERLAKADVRYRVIDIANTLDKAAAT 369

RESULT 19
US-08-976-063E-20
; Sequence 20, Application US/08976063E
; Patent No. 6524831
; GENERAL INFORMATION:
; APPLICANT: Steinbuechel, Alexander
; APPLICANT: Priefert, Horst
; APPLICANT: Rabenhorst, Jurgen
; TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF CONIFERYL
; FILE OF INVENTION: ALCOHOL, CONIFERYLALDEHYDE, FERULIC ACID, VANILLIN AND
; TITLE OF INVENTION: VANILLIC ACID AND THEIR USE
; FILE REFERENCE: Bayer-9998-CAO
; CURRENT APPLICATION NUMBER: US/08/976.063E
; PRIOR FILING DATE: 1997-11-21
; PRIOR FILING DATE: 1996 49 655.1 GERMANY
; PRIOR FILING DATE: 1996-11-29
; NUMBER OF SEQ ID NOS: 45


```

; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2152 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-036-987A-3

Query Match      6.0%; Score 115.5; DB 3; Length 2152;
Best Local Similarity 22.6%; Pred. No. 0.05;
Matches 79; Conservative 48; Mismatches 129; Indels 93; Gaps 19;

QY 17 AVKTDVSVPELKEGTALVKVEAVAGNPTDKHIAKYKIGPEGSLGCDIAGTVVKLGPN 136
Db 1419 ALVDEPTATAPLGDGEVRIAMRAAGVNFDRD-ALIALGMYPGVASLGSEGAGVVETGPGV 1477
QY 77 STDLKVGDTGFGFVHGASQTDPKNGAFABYARVYPPLFYKSNLTHSTADBISEGPKVNF 136
Db 1478 -TGLAPGDRVMGMI-----PK-----AFGPL---AVADHRMVTTRIPAG--WSFA 1515
QY 137 SAASLPVSLTTAGVSLCHHLGSKMEWHPTPQHTHPLLWGGATAVGQOLIQAOKHINAY 196
Db 1516 RAASVPIVFLTAYYALVDLAGLR-----PG-----ESLLVHSAAGGVGMAAIQIARHLGA- 1565
QY 197 TKIVTVASK-----KHEKLLKS-----YGADDVVF-----DYHDAGV- 227
Db 1566 -EYVATASEDKQAVELSHREHLASSRTCDPEQOFLGATGGRGVVDVVLNSLAGSFADASLR 1624
QY 228 -----IE-----QIKSKYPNLQH-VIDAV-GSEDSIPAYKVTADSLPATLL 267
Db 1625 MLPRGGRFLELTKTDVRDPVEVADAHGVSQAFDTVEAGPQRIGEMHLMHVELFEGRVL 1684
QY 268 EVVPMT---IESIPREIR-----KONVKIDITL--LYRASGOEILLGAT 306
Db 1685 EPLPVTAMDVRQAPALRHLSQLRHVKGKLVLTMPVWDAAAGTVLVTTGGT 1733

RESULT 25
US-09-603-207-3
; Sequence 3, Application US/09603207B
; Patent No. 6521406
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J
; APPLICANT: Turner, Jan R
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 DIV1
; CURRENT APPLICATION NUMBER: US/09/603, 207B
; CURRENT FILING DATE: 2000-06-23
; EARLIER APPLICATION NUMBER: 09/370, 700
; EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2152
; TYPE: PRT
; ORGANISM: Saccharopolyspora spinosa
; US-09-603-207-3

Query Match      6.0%; Score 115.5; DB 4; Length 2152;
Best Local Similarity 22.6%; Pred. No. 0.05;
Matches 79; Conservative 48; Mismatches 129; Indels 93; Gaps 19;

QY 17 AVKTDVSVPELKEGTALVKVEAVAGNPTDKHIAKYKIGPEGSLGCDIAGTVVKLGPN 136
Db 1419 ALVDEPTATAPLGDGEVRIAMRAAGVNFDRD-ALIALGMYPGVASLGSEGAGVVETGPGV 1477
QY 77 STDLKVGDTGFGFVHGASQTDPKNGAFABYARVYPPLFYKSNLTHSTADBISEGPKVNF 136
Db 1478 -TGLAPGDRVMGMI-----PK-----AFGPL---AVADHRMVTTRIPAG--WSFA 1515
QY 137 SAASLPVSLTTAGVSLCHHLGSKMEWHPTPQHTHPLLWGGATAVGQOLIQAOKHINAY 196
Db 1516 RAASVPIVFLTAYYALVDLAGLR-----PG-----ESLLVHSAAGGVGMAAIQIARHLGA- 1565
QY 197 TKIVTVASK-----KHEKLLKS-----YGADDVVF-----DYHDAGV- 227
Db 1566 -EYVATASEDKQAVELSHREHLASSRTCDPEQOFLGATGGRGVVDVVLNSLAGSFADASLR 1624
QY 228 -----IE-----QIKSKYPNLQH-VIDAV-GSEDSIPAYKVTADSLPATLL 267
Db 1625 MLPRGGRFLELTKTDVRDPVEVADAHGVSQAFDTVEAGPQRIGEMHLMHVELFEGRVL 1684
QY 268 EVVPMT---IESIPREIR-----KONVKIDITL--LYRASGOEILLGAT 306
Db 1685 EPLPVTAMDVRQAPALRHLSQLRHVKGKLVLTMPVWDAAAGTVLVTTGGT 1733

RESULT 24
US-09-370-700-3
; Sequence 3, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J
; APPLICANT: Turner, Jan R
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 DIV1
; CURRENT APPLICATION NUMBER: US/09/370, 700
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2152
; TYPE: PRT
; ORGANISM: Saccharopolyspora spinosa
; US-09-370-700-3

Query Match      6.0%; Score 115.5; DB 3; Length 2152;
Best Local Similarity 22.6%; Pred. No. 0.05;
Matches 79; Conservative 48; Mismatches 129; Indels 93; Gaps 19;

QY 17 AVKTDVSVPELKEGTALVKVEAVAGNPTDKHIAKYKIGPEGSLGCDIAGTVVKLGPN 136
Db 1419 ALVDEPTATAPLGDGEVRIAMRAAGVNFDRD-ALIALGMYPGVASLGSEGAGVVETGPGV 1477
```

Db 1685 EPLPVTAMDVRQAPALRHLQARHVGKLVLTMPVWDAAAGTIVLTGGT 1733

RESULT 26

US-08-804-227C-10
; Sequence 10, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only.
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3724 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-804-227C-10

Query Match 5.9%; Score 114; DB 2; Length 3724;

Best Local Similarity 23.7%; Pred. No. 0.17;
Matches 85; Conservative 48; Mismatches 139; Indels 86; Gaps 18;

QY 12 IEQKAVVKTDSVPELKEGTALVKVEAVAGNPTDKWHIAIKIGPEGSILGCDIAGTVVK 71
Db 2955 LDGLALVPAPDAEAP-LEPCQVRVAVRAAGVNFDR-ALIALGMYPGAEAMGTGAGTVVE 3012
QY 72 LGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAYARVYPPFYKSNLTHSTADEISEGP 131
Db 3013 VGPGV-TGVAVGDRVLGLWDGG-----LGPL-----CVADHRLAP 3047
QY 132 VK---NFESAASLPVSLTTAGVSLCHLGSKMWHPTPQHTHPLLIWGATAVGQOLIQ 188
Db 3048 VPDGWSYAQAASVPAPVFLSAYGLVTLAQLR-----PG-----ERVLVHAAAGGVGMAAVQ 3098
QY 189 VAKHINAYTKIVTVASKKHEKLLKSYGADDVFDYHDAGVIEQIKSKYPNLQHVDAVGSE 248
Db 3099 IARHLGA--EVLATASPGKWDALRAMGTD-----DHLASS 3132
QY 249 DSIPAEYKVT-ADSLPATLLEVVPMPTIESIPREIRKDNVKIDITL-LYRASGOEILLGAT 306
Db 3133 RTLDFAFAFTGADG--TSRADVV---LNSLTKEF-----VDASLGLRPGGRFLELGT 3181
QY 307 RFPASPEYHEATVKVFKFINPLNN-----GDIHNMNIKVFSNG-LDDVPALTGEGIK 357
Db 3182 DV-RDPERIAAEHPGVRYRAFDLNEAGPDALGRLLRELMDLFAAGVHLPLFPVVTHDVR 3238

RESULT 28
US-09-107-532A-5638

RESULT 27

US-08-804-198-4
; Sequence 4, Application US/08804198
; Patent No. 5945320
; GENERAL INFORMATION:
; APPLICANT: Burgett, Stanley G.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rao, Nagaraja R.
; APPLICANT: Richardson, Mark A.
; APPLICANT: Rostock, Paul R., Jr.
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: PAUL R. CANTRELL 1138
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,198
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CANTRELL, PAUL R.
; REGISTRATION NUMBER: 36,470
; REFERENCE/DOCKET NUMBER: P9113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3885
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3724 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-804-198-4

Query Match 5.9%; Score 114; DB 2; Length 3724;

Best Local Similarity 23.7%; Pred. No. 0.17;
Matches 85; Conservative 48; Mismatches 139; Indels 86; Gaps 18;

QY 12 IEQKAVVKTDSVPELKEGTALVKVEAVAGNPTDKWHIAIKIGPEGSILGCDIAGTVVK 71
Db 2955 LDGLALVPAPDAEAP-LEPCQVRVAVRAAGVNFDR-ALIALGMYPGAEAMGTGAGTVVE 3012
QY 72 LGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAYARVYPPFYKSNLTHSTADEISEGP 131
Db 3013 VGPGV-TGVAVGDRVLGLWDGG-----LGPL-----CVADHRLAP 3047
QY 132 VK---NFESAASLPVSLTTAGVSLCHLGSKMWHPTPQHTHPLLIWGATAVGQOLIQ 188
Db 3048 VPDGWSYAQAASVPAPVFLSAYGLVTLAQLR-----PG-----ERVLVHAAAGGVGMAAVQ 3098
QY 189 VAKHINAYTKIVTVASKKHEKLLKSYGADDVFDYHDAGVIEQIKSKYPNLQHVDAVGSE 248
Db 3099 IARHLGA--EVLATASPGKWDALRAMGTD-----DHLASS 3132
QY 249 DSIPAEYKVT-ADSLPATLLEVVPMPTIESIPREIRKDNVKIDITL-LYRASGOEILLGAT 306
Db 3133 RTLDFAFAFTGADG--TSRADVV---LNSLTKEF-----VDASLGLRPGGRFLELGT 3181
QY 307 RFPASPEYHEATVKVFKFINPLNN-----GDIHNMNIKVFSNG-LDDVPALTGEGIK 357
Db 3182 DV-RDPERIAAEHPGVRYRAFDLNEAGPDALGRLLRELMDLFAAGVHLPLFPVVTHDVR 3238

RESULT 28
US-09-107-532A-5638

```
; Sequence 5638, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5638:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 200 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (8) LOCATION 1...200
; SEQUENCE DESCRIPTION: SEQ ID NO: 5638:
US-09-107-532A-5638

Query Match      5.9%; Score 113.5; DB 4; Length 200;
Best Local Similarity 26.3%; Pred. No. 0.0018;
Matches 56; Conservative 37; Mismatches 75; Indels 45; Gaps 11;

Qy 13 EGDKAVKVDVSVELKEGTALVRVEAVAGNPTDVKHIAVKIGPEGIL-----GCD 64
Db 18 EAQDVFEEIDAHPREVGSGHVRVEIKAFSVNPDY---VALRLGEMKEIRTWKFPYVPGND 74

Qy 65 IAGTVVKGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAYARVYPLFYKSNLTHSTA 124
Db 75 GAGLVTIGSDVIT-VHVGOR--VAVHAV-----GGTYGE--KVVLTP-----S 112

Qy 125 DEISEGPKV-NFESAASLPVSLTAGVSLCHHLGSKMEWHPSTPQHTHPLLMGGATAG 183
Db 113 AKVAKIPDKMSWEAEAGVTPGITA-YNLNLH-----TEIQPTDVTWMLGASAVG 163

Qy 184 QQLIQVAKHNATYKIVTVASKKHEKLLKSYGA 216
Db 164 SSLIQLLHEKG--IRILTSASSKNEEKVTKFGA 194

RESULT 29
US-08-828-010-2
; Sequence 2, Application US/08828010
```

```
; Patent No. 5955355
; GENERAL INFORMATION:
; APPLICANT: Chen, Hong
; APPLICANT: Frielem, Nelson B.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: THE DIAGNOSIS AND TREATMENT OF NEUROPSYCHIATRIC
; TITLE OF INVENTION: DISORDERS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036/2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/828,010
; FILING DATE: 27-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-068
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 300 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; ORGANISM:
; FEATURE:
US-08-828-010-2
```

```
Query Match      5.8%; Score 111; DB 2; Length 300;
Best Local Similarity 24.0%; Pred. No. 0.0061;
Matches 80; Conservative 42; Mismatches 118; Indels 94; Gaps 17;

Qy 52 YKIGPEGSILGCDIAGTVVKGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAYARVYP 111
Db 17 FDIGFEG-----IGEVVALGLSASARYTVG-----QAVAYMAP--GSPAEYTVVPA 60

Qy 112 PLFYKSNLTHSTADEISEGPKVNF-SAASLPVSLTAGVSLCHHLGSKMEWHPSTPQHT 170
Db 61 SI-----ATPVPSEVKPEYLTLLVSGTTAVISL-KELGGLSE-----G 96

Qy 171 HPLLWGGATACQQLIQVAKHNATYKIVTVASKKHEKLLKSYGADDDVYHDAGVIEQ 230
Db 97 KKVLTAAAGGTGQPAQLSKAKACHV-IGTCSSEKSAFLKSLGCDRPNYKTEPVGTV 155

Qy 231 IKSYP-NLQHVTDVGSSEDSPEAYKVTAADSLPA-----TLLEVPWMTIE 275
Db 156 LKQYEGVDVVYVESVGG-----AMFDLAVDALATKRLIVTGFISGYQTPTGLSPVKAG 210

Qy 276 SIPEEIRKONVKIDITLL-----YPA-----SQO-----EILLGATRFPSPEYH- 315
Db 211 TLPKLLKKSASVOGFFLNHYLSKYQAAMSHLLMECVSGDLVCEVDLG-----DLSPEGRF 266

Qy 316 ----EATVKFVKFINPHLNNGDI-----HNMNIVK 341
Db 267 TGLSIFRAVNYMGMKNTGKIIVPELPHSVNSKL 300
```

Search completed: January 10, 2004, 10:08:46
Job time : 34 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2004, 10:07:24 ; Search time 71 Seconds
(without alignments)
1050.121 Million cell updates/sec

Title: US-10-081-644-2

Perfect score: 1920

Sequence: 1 MSVPTQKAVIEGKAVK.....ALTEGIKGNKNVYVARL 370

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747907 seqs, 201509753 residues

Total number of hits satisfying chosen parameters: 747907

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1920	100.0	370	14	US-10-081-644-2
2	1065.5	55.5	376	14	US-10-081-644-4
3	1021.5	53.2	377	14	US-10-081-644-6
4	1015.5	52.9	368	14	US-10-081-644-8
5	266	13.9	318	15	US-10-137-036-66
6	265	13.8	322	9	US-09-815-242-10703
7	221	11.5	331	15	US-10-156-761-8740
8	207	10.8	308	15	US-10-156-761-8739
9	206	10.7	338	9	US-09-815-242-10944
10	201.5	10.5	315	15	US-10-156-761-14981
11	197.5	10.3	329	12	US-10-369-493-16579
12	197	10.3	331	15	US-10-156-761-8218
13	193.5	10.1	334	15	US-10-156-761-9593
14	193	10.1	343	12	US-10-369-493-102
15	186.5	9.7	328	12	US-10-369-493-11505

Sequence 12902, A
Sequence 14785, A
Sequence 16435, A
Sequence 679, App
Sequence 6825, Ap
Sequence 8836, Ap
Sequence 3394, Ap
Sequence 5302, Ap
Sequence 65, Appl
Sequence 1893, Ap
Sequence 14009, A
Sequence 9450, Ap
Sequence 3730, Ap
Sequence 11902, A
Sequence 69, Appl
Sequence 53, Appl
Sequence 4692, Ap
Sequence 1245, Ap
Sequence 20316, A
Sequence 20590, A
Sequence 21665, A
Sequence 23606, A
Sequence 11926, A
Sequence 26, Appl
Sequence 6782, Ap
Sequence 677, App
Sequence 72, Appl
Sequence 12293, A
Sequence 1878, Ap
Sequence 5629, Ap
Sequence 12251, A
Sequence 2, Appl
Sequence 10, Appl
Sequence 1928, Ap
Sequence 23095, A
Sequence 4703, Ap
Sequence 7461, Ap
Sequence 8236, Ap
Sequence 2023, Ap
Sequence 20721, A
Sequence 1263, A
Sequence 5117, Ap
Sequence 14, Appl
Sequence 3785, Ap
Sequence 20, Appl
Sequence 20, Appl
Sequence 355, App
Sequence 10162, A
Sequence 833, App
Sequence 6285, Ap
Sequence 10436, A
Sequence 11395, A
Sequence 14789, A
Sequence 11244, A
Sequence 16582, A
Sequence 52, Appl
Sequence 53, Appl
Sequence 1472, Ap
Sequence 11164, A
Sequence 7741, Ap
Sequence 43, Appl
Sequence 45, Appl
Sequence 43, Appl
Sequence 45, Appl
Sequence 16399, A
Sequence 14953, A
Sequence 23041, A
Sequence 21022, A
Sequence 21322, A
Sequence 14207, A
Sequence 262, App
Sequence 23134, A

89 114 5.9 350 12 US-10-369-493-22735 Sequence 22735, A
90 114 5.9 1298 12 US-10-238-075-959 Sequence 959, App

ALIGNMENTS

RESULT 1

US-10-081-644-2
; Sequence 2, Application US/10081644
; Publication No. US20020192782A1
; GENERAL INFORMATION:
; APPLICANT: Yamamoto, Hiroaki
; APPLICANT: Kimoto, No. US20020192782A1hiro
; TITLE OF INVENTION: NOVEL ENONE REDUCTASES, METHODS FOR
; TITLE OF INVENTION: PRODUCING SAME, AND METHODS FOR SELECTIVELY REDUCING A
; TITLE OF INVENTION: CARBON-CARBON DOUBLE BOND OF AN ALPHA,BETA-UNSATURATED KETONE
; TITLE OF INVENTION: USING THE REDUCTASES
; FILE REFERENCE: 06501-100001
; CURRENT APPLICATION NUMBER: US/10/081,644
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: JP 2001-49363
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 370
; TYPE: PRT
; ORGANISM: Kluyveromyces lactis
US-10-081-644-2

Query Match 100.0%; Score 1920; DB 14; Length 370;
Best Local Similarity 100.0%; Pred. No. 3.7e-172;
Matches 370; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSVPTTQKAVIIEGDKAVVKTDSVPELKEGTALVKVEAVAGNPTDMKHIAIKIGPEGSI 60
DB 1 MSVPTTQKAVIIEGDKAVVKTDSVPELKEGTALVKVEAVAGNPTDMKHIAIKIGPEGSI 60
QY 61 LGCDIAGTVVVKLGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAYARVYVPLFYKSNLT 120
DB 61 LGCDIAGTVVVKLGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAYARVYVPLFYKSNLT 120
QY 121 HSTADEISEGPKNFESAASLPVSLTTAGVSLCHLGSKMEWHPSTPOHTHPLLIWGAT 180
DB 121 HSTADEISEGPKNFESAASLPVSLTTAGVSLCHLGSKMEWHPSTPOHTHPLLIWGAT 180
QY 181 AVGQOLIQAQKHINAYTKIVTVASKKHEKLLKSYGADDDVFDYHDAGVIEQIKSKYPNLQH 240
DB 181 AVGQOLIQAQKHINAYTKIVTVASKKHEKLLKSYGADDDVFDYHDAGVIEQIKSKYPNLQH 240
QY 241 VIDAVGSEDSIPEAYKVTADSLPATLLEVPMTIESIPEIRKDNVKIDITLLYRASGQE 300
DB 241 VIDAVGSEDSIPEAYKVTADSLPATLLEVPMTIESIPEIRKDNVKIDITLLYRASGQE 300
QY 301 ILGATFPASPPEYHEATVFKVFINPHLNGDIHNNIKVFNGLDDVDPALTEGIEKGK 360
DB 301 ILGATFPASPPEYHEATVFKVFINPHLNGDIHNNIKVFNGLDDVDPALTEGIEKGK 360
QY 361 KKNKVVYVARL 370
DB 361 KKNKVVYVARL 370

RESULT 2

US-10-081-644-4
; Sequence 4, Application US/10081644
; Publication No. US20020192782A1
; GENERAL INFORMATION:
; APPLICANT: Yamamoto, Hiroaki
; APPLICANT: Kimoto, No. US20020192782A1hiro
; TITLE OF INVENTION: NOVEL ENONE REDUCTASES, METHODS FOR
; TITLE OF INVENTION: PRODUCING SAME, AND METHODS FOR SELECTIVELY REDUCING A

; TITLE OF INVENTION: CARBON-CARBON DOUBLE BOND OF AN ALPHA,BETA-UNSATURATED KETONE
; FILE REFERENCE: 06501-100001
; CURRENT APPLICATION NUMBER: US/10/081,644
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: JP 2001-49363
; PRIOR FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 376
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-081-644-4

Query Match 55.5%; Score 1065.5; DB 14; Length 376;
Best Local Similarity 54.0%; Pred. No. 1.2e-91;
Matches 201; Conservative 59; Mismatches 109; Indels 3; Gaps 2;

QY 2 SVPTTQKAVIIEGDKAVVKTDSVPELKEGTALVKVEAVAGNPTDMKHIAIKIGPEGSI 61
DB 4 SIPETWKAVVIENGKAVVQDIPPELEEGFVLKTVAVAGNPTDMKHIDFKIGPQCALL 63
QY 62 GCDIAGTVVVKLGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAYARVYVPLFYK--SN 118
DB 64 GCDAAQGIQVKGPNVDAARFAIGDYIYGVHIGASVRFPSNGAFAYSAISSETAYKPAE 123
QY 119 LTHSTADEISEGPKNFESAASLPVSLTTAGVSLCHLGSKMEWHPSTPOHTHPLLIWG 178
DB 124 FRLCGKDKLPEGPKVKSLEGAVALPSVSLTTAGMLTHTSHFGLDMTWKPSKAQRDPILFWG 183
QY 179 ATAVGQOLIQAQKHINAYTKIVTVASKKHEKLLKSYGADDDVFDYHDAGVIEQIKSKYPNL 238
DB 184 ATAVGQMLQLAKLNGFSKIIVASRKHEKLEKYGADDELFDYHDADVIEQIKKYNKI 243
QY 239 QHVIDAVGSEDSIPEAYKVTADSLPATLLEVPMTIESIPEIRKDNVKIDITLLYRASG 298
DB 244 PYLVDCVSNTEITIQVYKCAADDLDATVVQLTVLTKEDIKESDRQNVSIETLLYLG 303
QY 299 QEILGATFPASPPEYHEATVFKVFINPHLNGDIHNNIKVFNGLDDVDPALTEGIEKE 358
DB 304 NDVPFGTFTLPADPEYKEAAIKFIRKINDGETHHIPVKVYKNGLDDIPQLLDDIKH 363
QY 359 GKNKVVYVARL 370
DB 364 GRNSGEKLVAVL 375

RESULT 3

US-10-081-644-6
; Sequence 6, Application US/10081644
; Publication No. US20020192782A1
; GENERAL INFORMATION:
; APPLICANT: Yamamoto, Hiroaki
; APPLICANT: Kimoto, No. US20020192782A1hiro
; TITLE OF INVENTION: NOVEL ENONE REDUCTASES, METHODS FOR
; TITLE OF INVENTION: PRODUCING SAME, AND METHODS FOR SELECTIVELY REDUCING A
; TITLE OF INVENTION: CARBON-CARBON DOUBLE BOND OF AN ALPHA,BETA-UNSATURATED KETONE
; FILE REFERENCE: 06501-100001
; CURRENT APPLICATION NUMBER: US/10/081,644
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: JP 2001-49363
; PRIOR FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 377
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-081-644-6

Query Match 53.2%; Score 1021.5; DB 14; Length 377;

Best Local Similarity 51.7%; Pred. No. 1.6e-87; Mismatches 193; Conservative 67; Mismatches 110; Indels 3; Gaps 2;

QY 1 MSVPTTQKAVIIEGDKAVKVTDSVPELKEGTALVKVEAVAGNPTDWHKHYKIGPESGI 60
DB 4 VAIPETMKAVIEDGKAVVKEGIPPELEBEGFVLKTLAVAGNPTDWAHIDIKIGPQSI 63
QY 61 LGCDDIAGTVVVLGPNAS--TDLKVGDTGFGFVHGASQTOPKNGAFAYARVYPLFYKS-- 117
DB 64 LGCDAAGQIVKLGPAVNPXDFSGIDYIGFIFHGSSVRFPFNGAFABEYSAISTVWAYKSN 123
QY 118 NLTHSTADISSEGPVKNFSAASLPVSLTTAGVSLCHHLGSKMEWHPSTQTHPLLIWG 177
DB 124 ELKFLGEDVLPAGVRSLEGVATIPVSLTTAGVLVLTYNLGLDKMEPSTPQKRGPIILWG 183
QY 178 GATVAGQQLIQVAKHINAYTKIVTVASKKHEKLLKSYGADDDVFDYHDAGVIEQIKSKYKN 237
DB 184 GATVAGQSLIQLANGLNGFTKIIVASRKHEKLLKEYGADELFDYHDDIDVVEQIKHKNYN 243
QY 238 LQHVDAVGSBDSIPAYKVTADSLPATLLEVPVMTIESIPEIRKDNVKIDITLLYRAS 297
DB 244 ISYLVDCVANQDTLQOYKCAADKQDATIVELKNLTVNENKVRNRRQNTVIDIRLYSIG 303
QY 298 GQIILLGATRPASPEYHEATVKFVFNPHLNGDIHMMNFKVFSNGLDDVPALTEGIGK 357
DB 304 GHEVPFGNITLPADSEARKAAIKFIKFNPKINDGQIRHVPVRYKNGLCVPHILKDIK 363
QY 358 EGNKKNVKVAVL 370
DB 364 YGNSEKULVAVL 376

RESULT 4
US-10-081-644-8
; Sequence 8, Application US/10081644
; Publication No. US20020192782A1
; GENERAL INFORMATION:
; APPLICANT: Yamamoto, Hiroaki
; TITLE OF INVENTION: NOVEL ENONE REDUCTASES, METHODS FOR
; TITLE OF INVENTION: PRODUCING SAME, AND METHODS FOR SELECTIVELY REDUCING A
; TITLE OF INVENTION: CARBON-CARBON DOUBLE BOND OF AN ALPHA, BETA-UNSATURATED KETONE
; TITLE OF INVENTION: USING THE REDUCTASES
; FILE REFERENCE: 06501-100001
; CURRENT APPLICATION NUMBER: US/10/081,644
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: JP 2001-49363
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-081-644-8

Query Match 52.9%; Score 1015.5; DB 14; Length 368;
Best Local Similarity 53.0%; Pred. No. 5.8e-87;
Matches 194; Conservative 60; Mismatches 109; Indels 3; Gaps 2;

QY 8 KAVIEGDKAVKVTDSVPELKEGTALVKVEAVAGNPTDWHKHYKIGPESGILGCDIAG 67
DB 2 KAVIEGDKAVKVEGPIPELEBEGFVLKTLAVAGNPTDWAHIDYKVGQSGILGCDIAG 61
QY 68 TVVKLGPNAS--TDLKVGDTGFGFVHGASQTOPKNGAFAYARVYPLFYKS--NLTHSTA 124
DB 62 QIVKLGPAVDPKDFSGIDYIGFIFHGSSVRFPFNGAFABEYSAISTVWAYKSNELFLGE 121
QY 125 DETSEGPVKNFSAASLPVSLTTAGVSLCHHLGSKMEWHPSTQTHPLLIWGATVAGQ 184
DB 122 DVLVAPGVRSLGAAIPVSLTTAGVLVLTYNLGLNKLMEPSTPQKRGPIILWGATVAGQ 181
QY 185 QLIQVAKHINAYTKIVTVASKKHEKLLKSYGADDDVFDYHDAGVIEQIKSKYPNLQHVDA 244

DB 182 SLIQLANKLNGFTKIIVASRKHEKLLKEYGADQLFDYHDDIDVVEQIKHKNYNSLYLDC 241
QY 245 VGSSEDSIPEAYKVTADSLPATLLEVPVMTIESIPEIRKDNVKIDITLLYRASGOEILG 304
DB 242 VANQNTLQOYKCAADKQDATIVELTNLTVNENKVRNRRQNTVIDIRLYSIGHEVPEP 301
QY 305 ATRFPASPEYHEATVKFVFNPHLNGDIHMMNFKVFSNGLDDVPALTEGIGKGNKVN 364
DB 302 GITFPADPEARAAATEFVKFNFNPKISDQIHHIPARVYKNGLYDVPRILEDIKIGNSGE 361
QY 365 KYVAVL 370
DB 362 KLVAVL 367

RESULT 5
US-10-137-036-66
; Sequence 66, Application US/10137036
; Publication No. US20030101478A1
; GENERAL INFORMATION:
; APPLICANT: Perera, Ranjan
; APPLICANT: Rice, Stephen
; APPLICANT: Bagleton, Clare
; APPLICANT: Lasham, Annette
; APPLICANT: Wood, Marion
; APPLICANT: Visser, Elizabeth
; TITLE OF INVENTION: Compositions and Methods for the
; TITLE OF INVENTION: Modification of Gene Expression
; FILE REFERENCE: 11000.1036c4
; CURRENT APPLICATION NUMBER: US/10/137,036
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: PCT/NZ 01/00115
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: U.S. No. US20030101478A1 09/724,624
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: U.S. No. US20030101478A1 09/598,401
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: PCT/NZ00/00018
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: U.S. No. US20030101478A1 60/146,591
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: U.S. No. US20030101478A1 09/276,599
; PRIOR FILING DATE: 1999-03-25
; NUMBER OF SEQ ID NOS: 143
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Eucalyptus grandis
US-10-137-036-66

Query Match 13.9%; Score 266; DB 15; Length 318;
Best Local Similarity 33.6%; Pred. No. 1.8e-16;
Matches 91; Conservative 31; Mismatches 93; Indels 56; Gaps 11;

QY 3 VPTTOKAVIIEGDKAVV---KTDVSVPELKEGTALVKVEAVAGNPTDWHKHYKIGPE 57
DB 7 IPTKMKAVYRHRGVDVANVLGDDPELKEGQVLVKVLAALNPIDTARVKVIGKLP 66
QY 58 G----SILGCDIAGTVVVLGPNASTDLKVGDTGFGFVHGASQTOPKNGAFAYARVYPL 113
DB 67 GFSLPVAPGVDLAGVVKVGVREV-ELKVGDEYVGFMPHAK---KDGTLAEYAAVEESP 121
QY 114 FYKSNLTHSTADEISEGPVK-NFESAASLPVSLTT-----AGVSLCHHLGSKMEWHP 165
DB 122 -----LALPKPKLRFGEAASLPVVIQTAYGGLERAGLS----- 154
QY 166 TPQHTHPLLIWGATVAGQQLIQVAKHINAYTKIVTVASKKHEKLLKSYGADDDVFDYHDA 225
DB 155 ---HGKSLVLVAGGAGVGTLLIQLAKVFGASRVAATSTGKLELLKSLGADLAIDYTKV 211
QY 226 GVIEQIKSKYPNLQHVDAVGSSEDSIPEAYK 256


```

; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8739
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-09-815-242-10944

Query Match      10.8%; Score 207; DB 15; Length 308;
Best Local Similarity 26.9%; Pred. No. 6.2e-11;
Matches 98; Conservative 43; Mismatches 111; Indels 112; Gaps 19;

QY   8 KAVIIE---GDKAVKTDVSVPELKEGTALVKVEAVAGNPTDWHKIAYKIGPEGIL--- 61
    |||:| | : : : | | | | | : : : | | | | | : : :
DB   2 RAWVEQMGSGPELVVEREVARPEPGLNEVLVRVHAAGVNPVDFKTRA-----SCALIIEWG 56

QY   62 ----GCDIAGTVKVLGNASTDLKVGDTCFGFVHGASQTDPKGAFAY-----ARYVPP 112
    |||:| | : : : | | | | | : : : | | | | | : : :
DB   57 EVPVAGNDVSGTVEAVPGVGGM-FRPGDE----VYGMPLFRQAGAYAEYVVAFAHLPAP 111

QY   113 LFYKSNLTSTADEISEGPVKNFESAASLPVSLTTAGVSLCHHLGSKMHPSTPQHPTH 172
    |||:| | : : : | | | | | : : : | | | | | : : :
DB   112 --KPNALTH-----VQAALPLAALTAMQALVDTAGVRA-----GER 146

QY   173 LIIWGATVAGQOLIQAQKHINAYTKIVTVASKHEKLKSYGADDDFYHDAGVIBOIK 232
    |||:| | : : : | | | | | : : : | | | | | : : :
DB   147 VLVHAAAGGVGHLAGVQAKARGAYV-IGTASAGKHD-LRLQLGADEVIDYRTV----- 197

QY   233 SKYPNLQHVDAVGSEDSIPKAYKTADSLPATILEVVPMIESPIEKRNKVIDITL 292
    |||:| | : : : | | | | | : : : | | | | | : : :
DB   198 -----REFDAVGVD-----VLDBGLG-----QNAERSLTV 224

QY   293 LYRASGOELL-GATRFPAPEYHEATVKVFKNFP-HLNNGDIHHNMKVFNSGLDDVP 350
    |||:| | : : : | | | | | : : : | | | | | : : :
DB   225 L-RPGRLLTLRGPDVADVGVCHVAVWML--VEPDHL-----GLREIA 266

QY   351 ALTE 354
    |||
DB   267 ALAE 270

RESULT 9
US-09-815-242-10944
; Sequence 10944, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23

; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8739
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-10944

Query Match      10.7%; Score 206; DB 9; Length 338;
Best Local Similarity 28.1%; Pred. No. 8.9e-11;
Matches 83; Conservative 39; Mismatches 87; Indels 86; Gaps 16;

QY   7 QKAVIIEGDKAVVKTDSVPELKEGTALVKVEAVAGNPTDWHK-----IAYKIGPEG 58
    |||:| | : : : | | | | | : : : | | | | | : : :
DB   12 QKELALE-----EVPLETIHDNDVLVKIIAASINPIDLKTKDGKVMILLNYQM---P 60

QY   59 SILGCDIAGTVKVLGNASTDLKVGDTCFGFVHGASQTDPKN--GAPAEYARV--YPPPLF 114
    |||:~| | : : : | | | | | : : : | | | | | : : :
DB   61 LIIGSDFAIGVIVSUGKKVQ-NPRLDGDAYGRV-----PKNRVGTTFABYIAVDQAQAVAM 112

QY   115 YKSNLTSTADEISEGPVKNFESAASLPVSLTTAGVSLCHHLGSKMHPSTPQHPTHPLL 174
    |||:| | : : : | | | | | : : : | | | | | : : :
DB   113 KPNULT-----FEAAAAILP-----VGLTSYQALHDMNVQPG-----QKVL 149

QY   175 IINGCATVAGQOLIQAQKHINAYTKIVTVASKHEKLKSYGADDDFYHDAGVIEQIKSK 234
    |||:| | : : : | | | | | : : : | | | | | : : :
DB   150 IQAGSGGGITAIQLAKLAGAY--VATTSSKNKEWQALGADEVIDYRTQN-FEEVLSD 206

QY   235 YPNLQHVDAVGSEDSIPKAYKTADSLPATILE-----VVP-----MTIESIPEE 280
    |||:| | : : : | | | | | : : : | | | | | : : :
DB   207 Y---DYVFTMG-----GTILEKAFVSXVXKPGKVTLSGIPNE 241

RESULT 10
US-10-156-761-14981
; Sequence 14981, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 14981
; LENGTH: 315
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-14981

Query Match      10.5%; Score 201.5; DB 15; Length 315;
Best Local Similarity 25.9%; Pred. No. 2.1e-10;
Matches 80; Conservative 49; Mismatches 125; Indels 55; Gaps 15;

QY   1 MSVPTTQKAV---IIEGDKAVVKTDSVPELKEGTALVKVEAVAGNPTDWHKIAYK--IG 55
    |||:| | : : : | | | | | : : : | | | | | : : :
DB   1 MSTVNTMRISQDVVLGGPEVLKEVETERPAPRPNEVLVRVRAAGVNPDTMKHRTGTFGLG 60

```



```
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 12902
; LENGTH: 353
; TYPE: PRP
; ORGANISM: Aspergillus nidulans
US-10-369-493-12902

Query Match          9.5%; Score 182; DB 12; Length 353;
Best Local Similarity 23.6%; Pred. No. 1.7e-08;
Matches 90; Conservative 57; Mismatches 149; Indels 86; Gaps 15;

QY 3 VPTOKAVIIEGDKAVVKTDSV-----PELKE--GTALVKVEAVAGNPTDKW---HIA 51
Db 9 IPPTMKSLVYSTQPSVISLSFTASAPSPPLPSYANQLLIKVTLSNPADHKVPQHAT 68
QY 52 YKIGPEGSL-----GCDIAGTVVVKLGNASTDLKVGDTGFGFVHGASQTDPKNGAFA 104
Db 69 IPTGGRTLICGLPASGLDFAGKIVSVHPKNRTEPEGQ-----LVVGLARPTGTG 124
QY 105 EYARVYPLPKSNLTHSTDAETISEGPKVNFESAASLPVSLTTAGVSLCHHLG-SKWEH 163
Db 125 EYI-----LADANDVAH-----LPEG--VSVDAAACLGAVRTAYOSLKNYLDLSKFE-G 171
QY 164 PSTPOHPTPLIIGGATVAVGQOLIQAKHINAYTKIVTVASKKHEKLLKSYGADVDVYH 223
Db 172 GSGR-----VFINGSGGCGVFALQIAKMLGC--QVTTCSGRNIELVRDLGADEIDYT 225
QY 224 DAGVIEQISKYPNLQHVDAVSGSEDSIPEAYKVTADSLPATLLEVVPMTIESIPEIRK 283
Db 226 TTNVTELTKGLVYDVHIG-----LPG----- 251
QY 284 DNKIDITLYRASGQBIIGATRPASPSPYHEATVKVFPINPLNNGDIHNMNKKVPS 343
Db 252 -NLVYAECHHFKPSGVMVQGA-----GSILTAFWRAITPRFLGGRRWFVPLMMA 301
QY 344 NGLDDVPALTEGIEKGKKNVK 365
Db 302 NSKEDLVEVGNLKEGKLRVK 323

RESULT 17
US-10-369-493-14785
; Sequence 14785, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 14785
```

```
; LENGTH: 333
; TYPE: PRP
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-14785

Query Match          9.5%; Score 181.5; DB 12; Length 333;
Best Local Similarity 28.3%; Pred. No. 1.8e-08;
Matches 84; Conservative 37; Mismatches 117; Indels 59; Gaps 12;

QY 8 KAVIEGDKAVK-----TDVSPELKEGTALVKVEAVAGNPTDKW-----HTAYKI 54
Db 2 KAFV--DKYKKKGPLADMPNPGVIGANDVLRHATAINLLDSKVRDGEFKFLFPLR- 58
QY 55 GPEGSTLGCIDIAGTVVVKLGNASTDLKVGDTGFGFVHGASQTDPKNGAFAEYARVYPLF 114
Db 59 --PPFILGHDLAGTVIRVGANV-RQFKTGDEF-----APRDRHVRGTFAEMIAV----- 105
QY 115 YKSNLTHSTDAETISEGPKVNFESAASLPVSLTTAGVSLCH-----HLGSKWEHHPSTPQHT 170
Db 106 -----DAADLALKPTSLSMQAASIPLVGLTAWQALLIEVGKVKSGK----- 147
QY 171 HPLLWGGATVAVGQOLIQAKHINAYTKIVTVASKKHEKLLKSYGADVDVYHDAGVIEQ 230
Db 148 --VFIQAGSGGVGTFAIQIAKHLGA--TVATTTSAANAELVKSLGADVVIDYKTQD-PFQ 202
QY 231 IKSYPNLQHVDAVSGSEDS---IPEAYKVTADSLPATLLEVVPMTIESIPEIRKD 284
Db 203 VLSGYDLVLSQDAKTLEKSLNVRPFGKLIISGPPDVAFARSLKLESAPAFACRON 259

RESULT 18
US-10-369-493-16435
; Sequence 16435, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 16435
; LENGTH: 276
; TYPE: PRP
; ORGANISM: Bacillus thuringiensis
US-10-369-493-16435

Query Match          9.3%; Score 179; DB 12; Length 276;
Best Local Similarity 26.2%; Pred. No. 2.3e-08;
Matches 85; Conservative 49; Mismatches 123; Indels 68; Gaps 17;

QY 60 ILGCDIAGTVVVKLGNASTDLKVGDTGFGFVHGASQTDPKNGAFAEYARVYPLFYSNL 119
Db 5 ILGNDFSGVIVKVGSGV-TRFKVGDALY-----ARPKNKIGCTFAEYIAIH----- 49
QY 120 THSTDAETISEGPKVNFESAASLPVSLTTAGVSL--CHLGSKWEHHPSTPQHTPLLI 175
Db 50 ----EDDIALKP-KNLSFEAASIPLVGLTSYQALHDIMHL-----QKQKILI 93
QY 176 WGGATVAVGQOLIQAKHINAYTKIVTVASKKHEKLLKSYGADVDVYHDAGVIEQISKY 235
Db 94 HAGSGVGTFAIQIAKIMGA--TVTTASEAGSLVKSLGADQIINYK-----TEFEEIL 147
QY 236 PNLQHVDAVSGSEDSIPEAYKVTADSLPATLLEVVPMTIESIPEIRKONKIDITLLYR 295
Db 148 KNYDAVFDITIG-TTLEKSFNIKSG--GNIVSVSGMPNARPKGKFGSGFFK---TLLFS 201
```


—

Db 100 ANTDMQAASLPLVALTA-----WQVLVETARLQKGOKVLIHAGSGGVCTVA 146
Qy 187 IQAKHINAYTKIYTVASKKHEKLLKSYGADDDY-----HDAGVIEQIKSKYPN 237
Db 147 IQAKHIGAF--VATTSTANVEWVKALGADLVIDYQQDFASVLHDYDV-----194
Qy 238 LQHVDAVGSSEDSIPEAYKV-----TADSLPAT-----LLEVVPMTIESIP 278
Db 195 ---VLSLGA-DVLEKSLKVLKPGQLISISGPTTVQFAREHGLSWPLQOVMLLSLIGIR 250
Qy 279 EEIRKONWKIDITLLYRASGOEI 301
Db 251 RKARKQDVSFAFVFM-RANGAQL 272

RESULT 27

US-10-156-761-9450
; Sequence 9450, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9450
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9450

Query Match 7.8%; Score 150; DB 15; Length 445;
Best Local Similarity 21.3%; Pred. No. 2.6e-05;
Matches 93; Conservative 52; Mismatches 142; Indels 150; Gaps 17;
Qy 1 MSVPTTOKAVIIIGD-----KAVVTKDVSVPKKEGTALVKVEAVAGN 44
Db 21 LPLPDSYRAITVHKDETETEMFAGLSTRDKPRKSHLDDVPVPELGPCEALVAVMAGSVNY 80
Qy 45 TD-WKHAYKIGPEG-----SILGCDIAGTVVKLGNA-----76
Db 81 NSVWTSIFEPVSTFNFLERYGRISDLSKRHDLPHYIIGSLAGVLTGTGVSNSWRPGDE 140
Qy 77 -----STDLKVGDTGFGFVHGASOTDPKN-----GAFAYEARVYPPPLFKSNLTH 121
Db 141 VVAHCLSVLESSDG-----HNDTMDPEORINGCFETNFGGLAEIALV-----KSNQLM 189
Qy 122 STADEISEGPVKNFESAASLPVSLTTAGVSLCHLHLSKMEWHPSTPQHTPLHGWGATA 181
Db 190 PKPDHLS-----WEEAAPGLVNSTAYRQLVSRNGAGMKQGDN-----VLIWGASGG 236
Qy 182 VQOQLIQVAKHINAYTKIYTVASKKHEKLLKSYGADDDYHDAGVIEQIKSKYPNLOHV 241
Db 237 LGSYATQFALAGGA-NPFCVVSSEQKADICRSMGAEAIIIDRN-----277
Qy 242 IDAVGSEDSIPEAYKVTADSLPATLLEVVPMTIESIPEEIRKONWKIDITLLYRASGOEI 301
Db 278 -----AEGYKFWKDETTQD-----PKEWKRFGKRIREFTGGEIDIVFEHFGRE- 321
Qy 302 LLGATRF-----PASPEYHEATVKFV-----KFINPHLNN-----331
Db 322 TFGASVYVTRKGGTITTCASTSGYMEYDNRYLWMSLKRIGSHFANYREAWEANRLVAK 381

Qy 332 GDIIHMNIKVFNSGLDD 348
Db 382 GKIHPTLSKVYS--LED 396
RESULT 28
US-09-738-626-3730
; Sequence 3730, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 3730
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3730

Query Match 7.8%; Score 149; DB 10; Length 318;
Best Local Similarity 27.0%; Pred. No. 1.9e-05;
Matches 67; Conservative 33; Mismatches 96; Indels 52; Gaps 11;
Qy 16 KAVVKT-----DVSVPKKEGTALVKVEAVAGNPTDWKHIA--YKIGPEG--I 60
Db 2 QAIQVTEEKVTASLELQEVFPVTLKPEGLVEVKAAGVNRADLLQTOGNYPV-PAGASEI 60
Qy 61 LGCDIAGTVVKLGPNASTDLKVGDTGFGFVHGASOTDPKNGAFAYEARVYPPPLFKSNLT 120
Db 61 LGLECAGVIVDAGDTGQT---VGQEVACLLTG-----GGVAVVAV-----P 99
Qy 121 HSTADEISEGPVKNFESAASLPVSLTTAGVSLCHLHLSKMEWHPSTPQHTPLHGWGAT 180
Db 100 EGQLMPEPEG--YSFVEAASIVEVACTVMSNIGMLAQKE-----DTFLHGGAG 148
Qy 181 AVQOQLIQVAKHINAYTKIYTVASKKHEKLLKSYGADDDYHDAGVIEQIKSKYPNLOH 240
Db 149 GIGTFAIQGKALGV-TVAVTAGSTEKTKCNLGLADILINYKEEDFAEVLKNNK---ADV 204
Qy 241 VIDAVGSE 248
Db 205 ILDIIGAK 212
RESULT 29
US-10-156-761-11902
; Sequence 11902, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI

```
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 11902
; LENGTH: 326
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-11902

Query Match          7.7%; Score 148.5; DB 15; Length 326;
Best Local Similarity 26.5%; Pred. No. 2.2e-05;
Matches 69; Conservative 37; Mismatches 99; Indels 55; Gaps 13;

Qy 14 GDKAVKTVDSVPELKEGTALVKEAVAGNPTD-WKHIAVKIGPEGS--ILGCDIAGTVV 70
Db 11 GPSALVMTVEPDPEPGESEVLVEVVASAVNRADLQRFYAPPPGASYPGLECSGRIA 70
Qy 71 KLGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAEYARVYPPPLFYKSNLTHSTADEISEG 130
Db 71 ALGPGVS-GWAVGDEVNALLAG-----GGVAEKVAV-----PAGQLLPV 108
Qy 131 PVK-NFESAASLP-VSLTT-AGVSLCHLGSKMWHHPSTPQHTHPLINGATAVGGQOLI 187
Db 109 PEGIGLRQAALPEVTCVNSVFMVAHL-----RPGET--LLVHGGSSGIGTMAI 157
Qy 188 QVAKHINAYTKIVTVASKKHEKLLKSYGADVDYHDGAVIEQIKSKYPNLQHVDAVGS 247
Db 158 QLAQAVGAKV-AVTAGTKEKLDFAELGADVLVNYREQDFVEVR-----BATDGAGA 209
Qy 248 EDSIPEAYKYVTADSLPATLL 267
Db 210 D-----VLDNMGAKYL 221

RESULT 30
US-09-908-744-69
; Sequence 69 Application US/09908744
; Publication No. US20030068791A1
; GENERAL INFORMATION:
; APPLICANT: Misaenikov Andrei
; APPLICANT: Ojamo Heikki
; APPLICANT: Povelainen Mira
; APPLICANT: Grob H+kan
; APPLICANT: Toivari Mervi
; APPLICANT: Richard Peter
; APPLICANT: Ruohonen Laura
; APPLICANT: Koivuranta Kari
; APPLICANT: Londeborough John
; APPLICANT: Aristidou Aristos
; APPLICANT: Penttil, Merja
; APPLICANT: Plazanet-Menut Claire
; APPLICANT: Deutscher Josef
; TITLE OF INVENTION: Manufacture of Five-Carbon Sugars and Sugar Alcohols
; FILE REFERENCE: 1427.0010005
; CURRENT APPLICATION NUMBER: US/09/908,744
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: PCT/FI01/000051
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: US 09/488,581
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 08/790,585
; PRIOR FILING DATE: 1997-01-29
; PRIOR APPLICATION NUMBER: US 08/368,395
; PRIOR FILING DATE: 1995-01-03

; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 11902
; LENGTH: 326
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-11902

Query Match          7.5%; Score 144.5; DB 11; Length 352;
Best Local Similarity 24.4%; Pred. No. 5.9e-05;
Matches 76; Conservative 35; Mismatches 109; Indels 91; Gaps 14;

Qy 22 DYSVPELKEGTALVKV--EAVAGNPTD-KHIAVKIGPEGSILGCDIAGTVV-KLGPNASTD 79
Db 22 DLSVPEATDDKVLIRKVAITYGICGSDIHTFKGEYKNPTTPVVLGHFSGQVVEVGANV-TK 80
Qy 80 LKVGDD-----TGFGFV-----HGASQTDPKNGAFAEYARVYPPPLFYKSN 118
Db 81 VKVGDRVTSETTF-YVCGECDYCKEYQYNLCPRKKGIGTQONGSMANY-----VLAREE 133
Qy 119 LTHSTADEISEGPKNFESAA-SLPVSLTTAGVSLCHLGSKMWHHPSTPQHTHPLING 177
Db 134 SIHLLPDHLS-----YEGAMSEPLACCVHMYQKSHLEK-----DTIIM 175
Qy 178 GATAVCQQLIQVAKHINAYTKIVTVASKKHE-KLLKSYGADVDYHD---AGVIBQIKS 233
Db 176 GPGPIGLYLQIATAKEIGAFVIMTGITKDAHRLALAKKLGADVIVDTMKEDLAKVNSITD 235
Qy 234 KYPNLQHVDAVGSSEDSIPEAYKYVTADSLPATLLEVVVPMTIESIPBEIRKDNVKIDITLL 293
Db 236 GY-GVDKVVDAAGA-----VPAVNASLP-----L 258
Qy 294 YRASGOEILLG 304
Db 259 IRKQGGFIQVG 269

Search completed: January 10, 2004, 10:16:57
Job time : 75 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2004, 08:39:46 ; Search time 74 Seconds
(without alignment)
793.633 Million cell updates/sec

Title: US-10-081-644-2

Perfect score: 1920

Sequence: 1 MSVPTTQKAVIEGDKAVK.....ALTEGIKGNKNVYVARL 370

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : A Geneseq 19Jun03:*

1: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1085.5	55.5	376	23	ABG93187
2	473	24.6	361	23	ABG93300
3	310.5	16.2	131	23	ABP10927
4	266.5	13.9	363	21	AA196748
5	266	13.9	318	21	AA27051
6	266	13.9	318	21	AA27051
7	265	13.8	322	22	AAU35110
8	264	13.8	313	23	AB47714
9	261	13.6	386	21	AAG43467
					S. cerevisiae BAX-
					C. albicans BAX-as
					Human ORFX protein
					A. terreus LovC de
					Auxin-induced prom
					Eucalyptus grandis
					Enterococcus faec
					Listeria monocytog
					Arabidopsis thalia

10	254	13.2	309	21	AAG43468	Arabidopsis thalia
11	254	13.2	386	21	AAG22152	Arabidopsis thalia
12	247	12.9	309	21	AAG22153	Arabidopsis thalia
13	206	10.7	338	22	AAU35351	Enterococcus faec
14	173	9.0	328	23	ABE54030	Lactococcus lactis
15	171	8.9	421	23	ABU10941	Yeast selected inc
16	168	8.8	337	22	AAG93071	C glutamicum prote
17	168	8.8	337	22	AAU79594	Corynebacterium gl
18	167	8.7	363	21	AAG28581	Arabidopsis thalia
19	167	8.7	363	21	AAG28581	Arabidopsis thalia
20	167	8.7	366	21	AAG28580	Arabidopsis thalia
21	167	8.7	366	21	AAG28580	Arabidopsis thalia
22	164.5	8.6	324	21	AAG39983	Arabidopsis thalia
23	164.5	8.6	329	23	AAG39982	Arabidopsis thalia
24	164.5	8.6	329	23	AB92914	Herbicidally activ
25	164.5	8.6	354	21	AAG39981	Arabidopsis thalia
26	162	8.4	262	21	AAG05068	Arabidopsis thalia
27	162	8.4	267	21	AAG05067	Arabidopsis thalia
28	162	8.4	292	21	AAG05066	Arabidopsis thalia
29	158.5	8.3	396	17	AAU97569	Interleukin-2 rece
30	158.5	8.3	413	22	ABU11836	Human IL-2R-associ
31	155	8.1	117	21	AAU27050	Auxin-induced prom
32	155	8.1	117	23	AAU80747	Eucalyptus grandis
33	155	8.1	334	22	AAU59488	Propionibacterium
34	153.5	8.0	336	16	AAU71935	Sec. alcohol-dehyd
35	151.5	7.9	340	23	ABP38413	Staphylococcus epi
36	151	7.9	325	21	AAU17585	Arabidopsis thalia
37	151	7.9	325	21	AAU50141	Arabidopsis thalia
38	149.5	7.8	322	20	AAU23994	Amino acid sequenc
39	149.5	7.8	332	21	AAU59181	Human oxidoreducta
40	149	7.8	318	22	AAG89976	C glutamicum prote
41	148.5	7.7	348	21	AAU17994	Arabidopsis thalia
42	148.5	7.7	348	21	AAU17994	Arabidopsis thalia
43	148.5	7.7	348	23	AB92696	Herbicidally activ
44	148.5	7.7	370	21	AAU27476	Arabidopsis thalia
45	147.5	7.7	351	21	AAG05685	Arabidopsis thalia
46	147.5	7.7	362	21	AAG05684	Arabidopsis thalia
47	147	7.7	343	21	ABP28623	Arabidopsis thalia
48	145.5	7.6	357	23	ABP28623	Streptococcus poly
49	144.5	7.5	352	22	AAU05993	Enterococcus avium
50	144.5	7.5	403	22	AAU61595	Propionibacterium
51	143	7.4	332	23	AB448452	Listeria monocytog
52	142.5	7.4	348	21	AAU82330	Pyrococcus horikos
53	142.5	7.4	367	23	ABU06016	Monascus purpureus
54	141.5	7.4	337	13	AAU36874	Alcohol dehydrogen
55	141.5	7.4	337	14	AAU35044	Insertion element
56	141.5	7.4	348	22	AAU96816	Putative P. abyssi
57	139.5	7.3	337	22	AAU36333	Pseudomonas aerugi
58	138.5	7.2	287	21	AAG28988	Arabidopsis thalia
59	138.5	7.2	291	21	AAG28987	Arabidopsis thalia
60	137.5	7.2	289	21	AAG20690	Arabidopsis thalia
61	137.5	7.2	308	21	AAG20689	Arabidopsis thalia
62	137.5	7.2	353	21	AAG20688	Arabidopsis thalia
63	137	7.1	308	21	AAU11238	Arabidopsis thalia
64	137	7.1	353	21	AAU11237	Arabidopsis thalia
65	137	7.1	356	21	AAU11236	Arabidopsis thalia
66	135.5	7.1	292	21	AAG05686	Arabidopsis thalia
67	134.5	7.0	306	22	AAG93028	C glutamicum prote
68	134.5	7.0	306	22	AAU79226	Corynebacterium gl
69	133.5	7.0	336	21	AAU95047	Candida albicans p
70	133	6.9	346	21	AAU04618	Arabidopsis thalia
71	132.5	6.9	396	17	AAU97570	Interleukin-2 rece
72	131.5	6.8	308	21	AAG47252	Arabidopsis thalia
73	131.5	6.8	353	21	AAG47251	Arabidopsis thalia
74	131.5	6.8	368	21	AAG47250	Arabidopsis thalia
75	131	6.8	333	22	AAU34133	Staphylococcus aur
76	131	6.8	335	22	AAU36558	Staphylococcus aur
77	131	6.8	363	20	AAU95109	Amino acid sequenc
78	131	6.8	377	22	AAU19931	Human oxidoreducta
79	131	6.8	377	23	AAU51614	Human Adhr-1 prote
80	130.5	6.8	340	23	AAE23267	Bacillus strain LN
81	130.5	6.8	377	23	ABP69791	Human polypeptide
82	129	6.7	334	23	ABU04705	Mycobacterium tube

83 129 6.7 343 22 AAU30322 Novel human secret
84 128 6.7 292 21 AAG04620 Arabidopsis thalia
85 128 6.7 307 21 AAG04619 Arabidopsis thalia
86 127.5 6.6 362 22 ABB66217 Drosophila melanog
87 127.5 6.6 362 22 ABB70568 Drosophila melanog
88 127 6.6 338 23 ABP30052 Streptococcus poly
89 126 6.6 2532 21 AAY96753 A. terreus LovF/Sc
90 125.5 6.5 308 21 AAG47268 Arabidopsis thalia

ALIGNMENTS

RESULT 1

ABG93187
ID ABG93187 standard; Protein; 376 AA.

AC ABG93187;

XX 21-NOV-2002 (first entry)

DT S. cerevisiae BAX-associated protein fragment SEQ ID 332.

DE Bax; Bax-resistance; cytostatic; fungicide; immunosuppressive; virucide;
XX vasotrophic; vaccine; gene therapy; proliferative disorder; cancer;
KW apoptosis; fungal; yeast; infection; autoimmune disease; ischaemia;
KW neurodegeneration; cell death.

XX Saccharomyces cerevisiae.

OS WO200264766-A2.

XX 22-AUG-2002.

XX 21-DEC-2001; 2001WO-EPI5398.

XX 22-DEC-2000; 2000EP-0870318.

XX 04-JAN-2001; 2001EP-0870002.

XX 09-JAN-2001; 2001EP-0870003.

XX (JANC) JANSSEN PHARM NV.

XX Contreras RH, Eberhardt I, Luyten WHML, Reekmans RJ;

XX WPI; 2002-667002/71.

XX N-PSDB; ABQ76453.

XX New isolated nucleic acid representing a synthetic Bax gene, useful as
PT medicament for treating, preventing and/or alleviating yeast or fungal
PT infections or proliferative disorders, or for preventing apoptosis in
PT certain diseases

PS Claim 36; Figure 1; 344pp; English.

XX This invention describes a novel nucleic acid representing a synthetic
XX Bax gene. The Bax gene of the invention is useful for identifying
XX Bax-resistant yeast or fungi, identifying, or obtaining and identifying
XX Candida spp. sequences that are differentially expressed in a pathway
XX eventually leading to programmed cell death or identifying inhibitors or
XX inhibitor sequences of Bax-induced cell death. The products of the
XX invention have cytostatic, fungicide; immunosuppressive, virucide and
XX vasotropic activity and can be used in vaccines or for gene therapy. The
XX isolated nucleic acids, polypeptides, pharmaceutical compositions,
XX antisense molecules and antibodies are useful as medicaments or in
XX preparing a medicament for treating, preventing and/or alleviating
XX diseases associated with yeast or fungi or proliferative disorders, such
XX as cancer, or for preventing apoptosis in certain diseases. The compounds
XX or polypeptides, or the genetically modified organism are useful for
XX preparing a medicament for modifying the endogenic flora of humans and
XX other mammals. The vaccine is useful for immunising against yeast or
XX fungal infections. Apoptosis-related diseases include autoimmune disease,
XX ischaemia, diseases related with viral infections or neurodegenerations.
XX This sequence represents a polypeptide associated with the Bax gene

CC described in the disclosure of the invention.

XX Sequence 376 AA;

XX Query Match 55.5%; Score 1065.5; DB 23; Length 376;

XX Best Local Similarity 54.0%; Pred. No. 1.4e-88;

XX Matches 201; Conservative 59; Mismatches 109; Indels 3; Gaps 2;

QY 2 SVPTTKAVIIEGDKAVKTDVSVPELKEGTALVKVEAVAGNPTDKHAYKIGPGSIL 61

DB 4 SIPETMKAVVIENGKAVVQDIPPELEBGFVLIKTVAVAGNPTDKHIDFKIGPGALL 63

QY 62 GCDIAGTVVVKLGPN-STDCLKVGTGTFVHGASQTDPKNGAPAEYARVVPPLFYK--SN 118

DB 64 GCDAAQIVKLGPNVDAARFAIGDYIYGVHGSVFPNGAPAEYASISSETAYKPARE 123

QY 119 LTHSTADEISEGPKVNFESAASLPVSLTTAGVSLCHHLGSKMEWHPSTPQHTHPLIWGG 178

DB 124 FRLCGKDKLPEGEVKSLEGAVALPVSLLTAGMLTHTSFGLDMTWKPSKAQRDQPIFLWGG 183

QY 179 ATAVGQOLIQVAKHINAYTKIVTVASKKHEKLLKSYGADDVFDHAGVIEQIKSKYPNL 238

DB 184 ATAVGQMLQLAKKLNGFSKIIIVASRKHKLKEYGADELFDYHDADVIEQIKKYNMI 243

QY 239 QHVIDAVGSEDSIPEAYKVTADSLPATLLEVVPMTIESIPERIKONKIDITLLYRASG 298

DB 244 PIVDCVSTETIQVYKCAADDLDATVVLTKEDIKEEDRRQNVSEGTLLYLG 303

QY 299 QEILLGATRPASPEYHEATVFKVPINPHLNGDIHHNMIKVFSNGLDDVPALTEGIKE 358

DB 304 NDVPFGTFTLPADPEYKEAAIKFIKPNKINDGEIHHIPVKVYKNGLLDIPQLDDIKH 363

QY 359 GKKNVYVARL 370

DB 364 GRNKGKLVAVL 375

RESULT 2

ID ABG93300 standard; Protein; 361 AA.

XX AC ABG93300;

XX 21-NOV-2002 (first entry)

XX C. albicans BAX-associated protein fragment SEQ ID 558.

XX Bax; Bax-resistance; cytostatic; fungicide; immunosuppressive; virucide;

XX vasotropic; vaccine; gene therapy; proliferative disorder; cancer;

XX apoptosis; fungal; yeast; infection; autoimmune disease; ischaemia;

XX neurodegeneration; cell death.

XX Candida albicans.

XX WO200264766-A2.

XX 22-AUG-2002.

XX 21-DEC-2001; 2001WO-EPI5398.

XX 22-DEC-2000; 2000EP-0870318.

XX 04-JAN-2001; 2001EP-0870002.

XX 09-JAN-2001; 2001EP-0870003.

XX (JANC) JANSSEN PHARM NV.

XX Contreras RH, Eberhardt I, Luyten WHML, Reekmans RJ;

XX WPI; 2002-667002/71.

XX N-PSDB; ABQ76566.

XX New isolated nucleic acid representing a synthetic Bax gene, useful as
PT medicament for treating, preventing and/or alleviating yeast or fungal

Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also useful for the identification of potential new targets for antibiotic development. The antisense nucleic acids can also be used to identify proteins used in proliferation, to express these proteins, and to obtain antibodies capable of binding to the expressed proteins. The proteins can be used to screen compounds in rational drug discovery programmes. The antisense nucleic acid sequence is also useful to screen for homologous nucleic acids which are required for cell proliferation in a wide variety of organisms. The present sequence represents an essential prokaryotic cellular proliferation protein.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/publ/pct](http://wipo.int/pub/publ/pct) sequences.

Dusserget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Coscart P;
 PI Daniels J, Goebel W, Krest J, Kuhn M, Ng E, Vazquez-Boland JA;
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
 PI Chakraborty T, Dommann E, Hain T, Berche P, Charbit A, Durant L;
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
 PI Madueno E, De Pablos B, Wenland J, Kaerst U, Entian K, Hauf J;
 PI Rose M, Voss H;
 DR WPI: 2002-010914/01.

Query Match	13.6%	Score 261,	DB 21;	Length 386;
Best Local Similarity	32.2%	Pred. No. 3.9e-15;		
Matches	88;	Conservative 37;	Mismatches 104;	Indels 44; Gaps 11;
QY	2	SVPTTKAVIEG----	DKAVKTDVSVPELKEGTALVKVEAVAGNPTDWRKIAYKI----	54
Db	73	SIPKEMKAWVSYDYGVDVVKLESNIWVPEIKEDQVLIKVAALNPVDKRRQGGFKAT	132	
QY	55	GPEGSILGDIAGTVVVKLGNASTLIKVDGTGFGVHGASQTDPKN-GAPAEYARVYPP	112	
Db	133	DSELPVTPGVYDAGVVKVVG-SAVKOLKEGDEVYANVSEKALEGPKQFGSLAEYTAVEEK	191	
QY	113	L--FYKSNLTHSTADEISEGPVKNPFSASLPSVLTATAGVSLCHLGSKMWHHPSTPQHT	170	
Db	192	LLALKPKNI-----DFQAAGLPJAIAETADEGLV-----RTEFSAG-----	227	
QY	171	HPILLINGGATAVGQOLITQAKHINAYTKIVTVASKKHEKLLKSGADDDVFDYHDAGVIRQ	230	
Db	228	KSILVLNGAGGVSLVQLAKHYGSKVAATASTKLELVRSLGADLAIDYTKEN-ID	286	
QY	231	IKSKYNLQHVIDAVGSEDS-----IPEAVKVTA	259	
Db	287	LPDKY---DVVFDIAGMCDKAVKVIKEGGKVA	316	

Db 124 -----DFAQAAGLPLAETADEGLV-----RTEFSAG-----KSILVLNAGGVSGLVI 167
Qy 188 QVAKHINAYTKIVTVASKKHEKLLKSYGADVDYHDAGVIEQIKSKYPNLQHVDAVGS 247
168 QLAHVYTGASKVAATASTEKLELVRSIGADLAIDYTKEN-IEDLPDKY----DVVFDAGM 223
Qy 248 EDS-----IPEAYKVTA 259
Db 224 CDRKAVKIKGGKVA 239

RESULT 11
AAG22152
ID AAG22152 standard; Protein; 386 AA.
AC AAG22152;
XX
XX 17-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment. SEQ ID NO: 24972.
DE Arabidopsis thaliana
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX
XX EF1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 07-MAY-1999; 99US-0132487.
XX 11-MAY-1999; 99US-0132863.
XX 14-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.


```
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151085.
PR 27-AUG-1999; 99US-0151086.
PR 27-AUG-1999; 99US-0151086.
PR 27-AUG-1999; 99US-0151086.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159291.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match 13 24; Score 254; DB 21; Length 386;
Best Local Similarity 31.5%; Pred. No. 1.7e-14;
Matches 86; Conservative 37; Mismatches 106; Indels 44; Gaps 11;
2 SVPTTQKAVIIEG----DKAVVKTDSVPPELKEGTALVKVEAVAGNPTDWKHIAKYI---54

Db 73 SIPKEMKAWYSDYGGVDVLKLESNTAVPEIKEDQVLIKVAAAGLPVDAKRRQGRPKAT 132
Qy 55 -GPEGILGCDIAGTVVVLKGPNASTDOLKVGDTGFGFVHGASQTDPKN-GAFAEYARVYPP 112
Db 133 DSELPVPGVDVAGVVKVG-SAVKDFKEGDEVYANVSEKALEGPKQFGSLABYTAVEEK 191
Qy 113 L--FYKSNLTHSTADBISEGPVKNFESAASLPVSLTTAGVSLCHILGSKNEMHPSTPQHT 170
Db 192 LLALKPKNI-----DFAQAAGLPLAETADEGLV-----RTEPSAG----- 227
Qy 171 HPLLINGGATAVCQQLIOVAKHINAYTKIVTVASKKHEKLLKSYGADDVDFYHDAGVIEQ 230
Db 228 KSLVLNGAGGVGLMIQLAKHYVGASKVAATASTGKLELVRLSGADLAIDYTKEN-IED 286
Qy 231 IKSKYPNLQHVIDAVGSEDS---IPEAVKYTA 259
Db 287 LPDKY---DVVFDAGMCDKAVKVIKEGGKVA 316

RESULT 12
AAG22153
ID AAG22153 standard; Protein; 309 AA.
AC AAG22153;
XX
DT 17-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 24973.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
```

PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139751.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159299.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160778.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.

PT New nucleotide sequence useful in the identification or Lactococcus
PT lactis and related species -
PS Claim 6; SEQ ID No 732; 2504pp; French.
XX
CC The present invention is related to a Lactococcus lactis nucleotide
CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The
CC nucleic acid sequence is useful in the detection and/or amplification of
CC nucleic acid sequence, particularly to identify Lactococcus lactis or
CC related species. The proteins of the invention are useful for the
CC biosynthesis or biodegradation of a composition of interest. The
CC invention helps research in lactic bacteria, particularly useful in the
CC production of yogurt and cheese.
CC Note: The sequence data for this patent is based on equivalent patent
CC WO2001/77334 (published 18-OCT-2001) which is available in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 328 AA;
Query Match 9.0%; Score 173; DB 23; Length 328;
Best Local Similarity 28.8%; Pred. No. 3.3e-07;
Matches 77; Conservative 43; Mismatches 91; Indels 44; Gaps 11;
QY 8 KAV-IIEGDKAVKTDVSPPELKEGTALVKVEAVAGNPTDWKHIAVKIGPEGS-----IL 61
DB 2 KAIGTIDSKNFIDFEIEKPLRPHDLLIKVEAISINPVDTK---VRKGIKGNLAEPKIL 58
QY 62 GCDIAGTVVKLPNASTDLKVGDTGP--GFVHGASQTDPKNGFAFAEVARVYPL--FYKS 117
DB 59 GWDGLGTVVELGSETKL-FKVGDKVFWAGDV-----TRSGNAEFQAVDERIVGFAPK 110
QY 118 NLTHSTADEISEGPKNFESAASI.PVSLTTAGVSLCHLGLSKMEWHPSTPQHTHPLLLMG 177
DB 111 NLTK-----EKAVAPLTSLAYELLFEKL-----EVTDKSKGKSLIIN 150
QY 178 GATAVGQOLIQAQKHINAYTKIIVTVASKHEKLKSYGADDDVFDYHDAGV--IEQIKSKY 235
DB 151 GAGGVGSVAIQAKNA-GLTVIATASNPAIEWVKNGFNAGDVTYNHHEKLVPOVHELGFKF 209
QY 236 PNQHVUIDAV 245
DB 210 VDYLILNAV 219
RESULT 15
ABJ10941
ID ABJ10941 standard; Protein; 421 AA.
AC ABJ10941;
XX
DT 10-DEC-2002 (first entry)
XX
DE Yeast selected interacting domain protein SEQ ID NO: 16.
XX
KW Yeast; protein-protein interaction; Selected Interacting Domain;
KW STD (RTM); secretion yield; cancer; neurodegenerative disease; fungicide;
KW cytostatic; neuroprotective.
XX
OS Saccharomyces cerevisiae.
XX
PN WO20026504-A2.
XX
PD 29-AUG-2002.
XX
PF 14-FEB-2002; 2002WO-EP02299.
XX
PR 16-FEB-2001; 2001US-269266P.
XX
PA (HYBR-) HYBRIGENICS.
XX
PI Legrain P;
XX
DR WPI; 2002-674913/72.

DR N-PSDB; ABT11258.
XX
PT New protein-protein complexes of Saccharomyces cerevisiae, useful in
PT drug screening or development, for developing yeast strains with better
PT secretion yield of protein, or in gene therapy (e.g. to treat Candida
PT infection or cancer) -
XX
PS Claim 6; Page 41; 357pp; English.
XX
CC The present invention relates to complexes between Saccharomyces
CC cerevisiae Selected Interacting Domain (SID (RTM)) proteins and coding
CC sequences. The protein complexes of S. cerevisiae are useful in drug
CC development, in screening drugs or agents that modulate the interaction
CC of proteins, for developing yeast strains with better secretion yield of
CC protein, and in gene therapy. The protein complexes, polypeptides and
CC polynucleotides are useful for preventing or treating Candida infection,
CC cancer or neurodegenerative diseases in humans or animals. The present
CC sequence is a protein of the invention.
XX
SQ Sequence 421 AA;
Query Match 8.9%; Score 171; DB 23; Length 421;
Best Local Similarity 46.5%; Pred. No. 7.3e-07;
Matches 33; Conservative 12; Mismatches 26; Indels 0; Gaps 0;
QY 300 EILLGATREFPASPEYHEATVKFKFNPILNNGDIHHMNIKVFSNGLDDVDPALTEGKEG 359
DB 2 EVFPGNITUPADSEAKRAAIKFIKIPNKINDGQIRHPIRVYKNGLCDVPHILKDIKYG 61
QY 360 KKNKVKYVARL 370
DB 62 KNSGEKLVAVL 72
RESULT 16
AAG93071
ID AAG93071 standard; Protein; 337 AA.
XX
AC AAG93071;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ ID NO: 6825.
XX
KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.
XX
OS Corynebacterium glutamicum.
XX
PN EPI108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR N-PSDB; AAH68290.
XX
PT Novel polynucleotides derived from Corynebacterium, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT -
XX
PS Claim 17; SEQ ID NO: 6825; 246pp + Sequence Listing; English.

PR	04-JUN-1999;	99US-0137502.	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	99US-0138094.	PR	03-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138540.	PR	03-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	99US-0138847.	PR	10-AUG-1999;	99US-0148171.
PR	14-JUN-1999;	99US-0139119.	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	99US-0139452.	PR	12-AUG-1999;	99US-0148341.
PR	16-JUN-1999;	99US-0139453.	PR	13-AUG-1999;	99US-0148565.
PR	17-JUN-1999;	99US-0139492.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139454.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139455.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139456.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139457.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151066.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0139889.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0140354.	PR	01-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	99US-0140695.	PR	07-SEP-1999;	99US-0152363.
PR	24-JUN-1999;	99US-0140823.	PR	10-SEP-1999;	99US-0153070.
PR	28-JUN-1999;	99US-0140991.	PR	13-SEP-1999;	99US-0153758.
PR	29-JUN-1999;	99US-0141099.	PR	15-SEP-1999;	99US-0154018.
PR	30-JUN-1999;	99US-0141287.	PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	99US-0141842.	PR	20-SEP-1999;	99US-0154779.
PR	01-JUL-1999;	99US-0142154.	PR	22-SEP-1999;	99US-0155139.
PR	02-JUL-1999;	99US-0142055.	PR	23-SEP-1999;	99US-0155486.
PR	06-JUL-1999;	99US-0142330.	PR	24-SEP-1999;	99US-0155659.
PR	08-JUL-1999;	99US-0142803.	PR	28-SEP-1999;	99US-0156458.
PR	09-JUL-1999;	99US-0142920.	PR	29-SEP-1999;	99US-0156596.
PR	12-JUL-1999;	99US-0142977.	PR	04-OCT-1999;	99US-0157117.
PR	13-JUL-1999;	99US-0143542.	PR	05-OCT-1999;	99US-0157753.
PR	14-JUL-1999;	99US-0143624.	PR	08-OCT-1999;	99US-0157865.
PR	15-JUL-1999;	99US-0144005.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144085.	PR	08-OCT-1999;	99US-0158232.
PR	16-JUL-1999;	99US-0144086.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144325.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144333.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144884.	PR	18-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144884.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145132.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160989.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145919.	PR	25-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146388.	PR	28-OCT-1999;	99US-0161920.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161992.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147204.	PR	29-OCT-1999;	99US-0162142.
PR	04-AUG-1999;	99US-0147302.			
PR	05-AUG-1999;	99US-0147192.			
PR	05-AUG-1999;	99US-0147260.			

Query Match 8.7%; Score 167; DB 21; Length 363;
Best Local Similarity 26.1%; Pred. No. 1.4e-06;

RESULT 21


```
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0162142.
PR 28-OCT-1999; 99US-0162142.

Query Match      8.7%; Score 167; DB 21; Length 366;
Best Local Similarity 26.1%; Pred. No. 1.4e-06;
Matches 72; Conservative 47; Mismatches 91; Indels 66; Gaps 15;

QY 2 SVPTTQKAVITE--GDKAV--VKTDVSVPELKEGTALVKVENVAGNPTDMK----- 48
Db 27 SVFTGCKAVILPRFGGEVFEFLRENVPVFNLPNEVLVKAKAVSNPLDCRIRAGYGRSV 86
QY 49 ---HIAYKIGPEGILGCDIAGTVVVLGPNASTDLKVGDTGTFGVHGSQTDPKNGAFAE 105
Db 87 FQPHLPI-----IVGRDVSGEVAAIGTSVKSLKVGQEVFGALHPTA----LRGTYYD 134
QY 106 YARYVPLFYKSNLTHSTADISBGP--VKNFESAASLPVSLTTPAGVSLCHILGSKNMEWH 163
Db 135 YGIL-----SEDELTEKPSISHSVE--ASAIPFAALTAMRAL----- 169
QY 164 PSTQHT--HPLLWGGATAGQOLIQVAKHINAYTKIVTVASKKHEKLKSKYGADVD 221
Db 170 KSNARITEGQRLLVFGGGGAVGFSAIQIYAVASGCHVTASCVGQTK--DRILAA--GAEQAVD 227
QY 222 YHDAGVTEQIKSKYPNLQHVLDVAGSEDSIPEAYKV 257
Db 228 YTTEDIELAVKGP---DAVLDTIGG-----PETERI 256

RESULT 22
AAG39983
ID AAG39983 standard; Protein; 324 AA.
AC AAG39983;
XX
XX
DT 18-OCT-2000 (first entry)
XX
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 49547.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0131825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 05-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 23-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139763.
XX 21-JUN-1999; 99US-0139817.
XX 22-JUN-1999; 99US-0139899.
XX 23-JUN-1999; 99US-0140353.
XX 23-JUN-1999; 99US-0140354.
XX 24-JUN-1999; 99US-0140695.
XX 28-JUN-1999; 99US-0140823.
XX 29-JUN-1999; 99US-0140991.
XX 30-JUN-1999; 99US-0141287.
XX 01-JUL-1999; 99US-0141842.
XX 01-JUL-1999; 99US-0142154.
XX 02-JUL-1999; 99US-0142055.
XX 06-JUL-1999; 99US-0142390.
XX 08-JUL-1999; 99US-0142803.
XX 09-JUL-1999; 99US-0142920.
XX 12-JUL-1999; 99US-0142977.
XX 13-JUL-1999; 99US-0143542.
XX 14-JUL-1999; 99US-0143624.
XX 15-JUL-1999; 99US-0144005.
XX 16-JUL-1999; 99US-0144085.
XX 16-JUL-1999; 99US-0144086.
XX 19-JUL-1999; 99US-0144325.
XX 19-JUL-1999; 99US-0144331.
XX 19-JUL-1999; 99US-0144332.
XX 19-JUL-1999; 99US-0144333.
XX 19-JUL-1999; 99US-0144334.
```

[illegible]

PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159330.	QY	14	GDKAVVKTDSVP	ELKEGTALVKVEAVAGNPTDKHIA	YKIGP-----EGSILCCDIAGT	68
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159331.	DB	12	GAAGLEHVQVP	PTPKSNEVCLKLEATSLNPVDWKI	QKGMIRPFLPRKFFCIPATDVAGE	71
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159637.	QY	69	VVKLGPNASTDL	KVGTGTFGVHGHASQTDPKNGAF	AEVARYVPLFYKSNLTHSTADEIS	128
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159638.	DB	72	VVEVGSVK-NP	KAGDKVAVLSHLG-----GGGLAEFA	-----VATEKLTVKRPOEVG	119
PR	21-JUL-1999;	99US-0144814.	PR	14-OCT-1999;	99US-0159638.	QY	129	EGPVKNFESAAS	LPVSLTTAGVSLCHHLGSKMEWH	PSPTQHTHPLLIWGATAVGOOLIQ	188
PR	21-JUL-1999;	99US-0144814.	PR	14-OCT-1999;	99US-0159584.	DB	120	AA-----EAA	ALPVAGLTALQALTNPAGLKLD--	-GTGKKAN-ILVTAASGGVGHVAVQ	169
PR	21-JUL-1999;	99US-0145086.	PR	14-OCT-1999;	99US-0160741.	QY	189	VAKHINAYTKI	VTVASKKHEKLLKSYGADDDVPDY	---HDAGVIBQIKSKY-----	235
PR	21-JUL-1999;	99US-0145086.	PR	14-OCT-1999;	99US-0160767.	DB	170	LAKLANAH--	VTATCGARNIEBFVKSGLGADE	LDYKTPGGAALKSPSGKKYDAVVHCANGI	227
PR	21-JUL-1999;	99US-0145088.	PR	14-OCT-1999;	99US-0160767.	QY	236	-----PNLOH-	---VIDAVGSEDSIPEAYKVTAD	SLPATLLEVVPMPTIESIPEIRKDNV	286
PR	22-JUL-1999;	99US-0145085.	PR	14-OCT-1999;	99US-0160768.	DB	228	PFSVPEPNLSE	NGKVIDITPGFNAM---WTYAVK	KITMSKQLVPLLL--IP---KAENL	279
PR	22-JUL-1999;	99US-0145087.	PR	14-OCT-1999;	99US-0160770.	QY	287	KIDITLLYRAS	QOEILLGATRPASPPEYHEATV	KVFPINPH 328	
PR	22-JUL-1999;	99US-0145192.	PR	14-OCT-1999;	99US-0160815.	DB	280	EFWNLVKEGK	VTVI--DSKHPLS-----KAED	AWAKSIDGH 315	
PR	23-JUL-1999;	99US-0145145.	PR	14-OCT-1999;	99US-0160980.	RESULT 23					
PR	23-JUL-1999;	99US-0145218.	PR	14-OCT-1999;	99US-0160981.	AAG39982					
PR	23-JUL-1999;	99US-0145224.	PR	14-OCT-1999;	99US-0160989.	ID	AAG39982	standard; Protein; 329 AA.			
PR	26-JUL-1999;	99US-0145276.	PR	14-OCT-1999;	99US-0161404.	XX	AAG39982;				
PR	27-JUL-1999;	99US-0145913.	PR	14-OCT-1999;	99US-0161405.	XX					
PR	27-JUL-1999;	99US-0145918.	PR	14-OCT-1999;	99US-0161406.	XX					
PR	28-JUL-1999;	99US-0145919.	PR	14-OCT-1999;	99US-0161359.	XX					
PR	02-AUG-1999;	99US-0146386.	PR	14-OCT-1999;	99US-0161360.	XX					
PR	02-AUG-1999;	99US-0146388.	PR	14-OCT-1999;	99US-0161361.	XX					
PR	02-AUG-1999;	99US-0146389.	PR	14-OCT-1999;	99US-0161920.	XX					
PR	03-AUG-1999;	99US-0147038.	PR	14-OCT-1999;	99US-0161992.	XX					
PR	04-AUG-1999;	99US-0147204.	PR	14-OCT-1999;	99US-0161993.	XX					
PR	04-AUG-1999;	99US-0147302.	PR	14-OCT-1999;	99US-0162142.	XX					
PR	05-AUG-1999;	99US-0147192.	PR	14-OCT-1999;	99US-0162142.	XX					
PR	05-AUG-1999;	99US-0147260.	PR	14-OCT-1999;	99US-0162142.	XX					
PR	06-AUG-1999;	99US-0147303.	PR	14-OCT-1999;	99US-0162142.	XX					
PR	06-AUG-1999;	99US-0147416.	PR	14-OCT-1999;	99US-0162142.	XX					
PR	09-AUG-1999;	99US-0147493.	PR	14-OCT-1999;	99US-0162142.	XX					
PR	09-AUG-1999;	99US-0147935.	PR	14-OCT-1999;	99US-0162142.	XX					
PR	10-AUG-1999;	99US-0148171.	PR	14-OCT-1999;	99US-0162142.	XX					
PR	11-AUG-1999;	99US-0148319.	PR	14-OCT-1999;	99US-0162142.	XX					
PR	12-AUG-1999;	99US-0148341.	PR	14-OCT-1999;	99US-0162142.	XX					
PR	13-AUG-1999;	99US-0148565.									

XX AAG39981;
 XX AC
 XX DT 18-OCT-2000 (first entry)
 XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 49545.
 XX KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX OS Arabidopsis thaliana.
 XX PN EPI033405-A2.
 XX PD 06-SEP-2000.
 XX PF 25-FEB-2000; 2000EP-0301439.
 XX 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 20-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 10-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.


```
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      8.6%; Score 164.5; DB 21; Length 354;
Best Local Similarity 23.4%; Pred. No. 2.2e-06;
Matches 80; Conservative 56; Mismatches 141; Indels 65; Gaps 16;

QY 14 GDKAVKTDVSVPELKEGTALVKVEAVAGNPTDWKHIAKIGP-----EGSILGCDIAGT 68
DB 42 GAAGLEHVQVPVTPKSEVCLKEATSLNPVDNKIQGMIRPFLPRKFCIPATDVAGE 101
QY 69 VVKLGPNASTDLKVGDTGFGFVHGASQTDPKNGFAFAEYARYVPLFYKSNLTHSTADEIS 128
DB 102 VVEVGSVK-NFRAGDKVAVLSHLG-----GGGLAEFA-----VATEKLTVRPOEVG 149
QY 129 EGPVKNFESAASLVSLTTAGVSLCHHLGSMENHPSTPQHTPLLIWGGATAVGQQLIQ 188
DB 150 AA-----EAAALFVAGLTALQALTNPAGLKLD---GTGKKAN-ILVTAASGGVGHYAVQ 199
QY 189 VAKHINAYTKVTIVASKKHEKLLKSYGADVDYD---HDAGVIBQIKSYK-----235
DB 200 LAKLANAH--VTATCGARNTEFVKSLGADEVLDTKTPGGAALKSPGKKYDAVVHCANGI 257
```

```
QY 236 -----PNLQH-----VIDAVGSEDSIPEAYKVTADSLPATLLEVVPMTIESIPEEIRKDNV 286
DB 258 PFSVPFNPENLSENGKVIDITPGPNAM---WTYAVKKITMSKKQLVPLLL--IP---KAENL 309
QY 287 KIDITILYRASGOEILLGATRPFPASPEYHEATVKFKFINPH 328
DB 310 EFMVNLVKEGKVKTVI--DSKHPLS----KAEDAWAKSIDGH 345

RESULT 26
AAG05068
ID AAG05068 standard; Protein; 262 AA.
XX AC AAG05068;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 1312.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 28-APR-1999; 99US-0130891.
XX PR 30-APR-1999; 99US-0131449.
XX PR 30-APR-1999; 99US-0132048.
XX PR 30-APR-1999; 99US-0132407.
XX PR 04-MAY-1999; 99US-0132484.
XX PR 05-MAY-1999; 99US-0132485.
XX PR 06-MAY-1999; 99US-0132486.
XX PR 06-MAY-1999; 99US-0132487.
XX PR 07-MAY-1999; 99US-0132863.
XX PR 11-MAY-1999; 99US-0134256.
XX PR 14-MAY-1999; 99US-0134218.
XX PR 14-MAY-1999; 99US-0134219.
XX PR 14-MAY-1999; 99US-0134221.
XX PR 14-MAY-1999; 99US-0134370.
XX PR 18-MAY-1999; 99US-0134768.
XX PR 19-MAY-1999; 99US-0134941.
XX PR 20-MAY-1999; 99US-0135124.
XX PR 21-MAY-1999; 99US-0135353.
XX PR 24-MAY-1999; 99US-0135629.
XX PR 25-MAY-1999; 99US-0136021.
XX PR 27-MAY-1999; 99US-0136392.
XX PR 28-MAY-1999; 99US-0136782.
XX PR 01-JUN-1999; 99US-0137222.
XX PR 03-JUN-1999; 99US-0137528.
XX PR 04-JUN-1999; 99US-0137502.
XX PR 07-JUN-1999; 99US-0137724.
XX PR 08-JUN-1999; 99US-0138094.
XX PR 10-JUN-1999; 99US-0138540.
```


Db 12 GAAGLEHVQVPPTPKSNEVCLKLEATSLNPVDWKIQKGMIRPFLPRKPPCIPATDVAGE 71
Qy 69 VVKLGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAYRYVYPLFYKSNLTHSTADEIS 128
Dy 72 VVQVSGVK-NFKAGDKVAVLSHLG-----GGGLAEFA-----VATEKLTVKRPQEVG 119
Qy 129 EGPVNFESAASLPVSLTTAGVSLCHHILGSKNEWHSPTPQHTHLLIWWGATAVGQOLIQ 188
Dy 120 AA-----EAAALPVAGLTALQALTNPAGLKLD---GTGKKAN-ILVTAASGGVGHVAVQ 169
Qy 189 VAKHNAVTKVTIVASKKHEKLLKSYGADDVFDY---HDAGVIEQIKSKYPNLOHVDAV 245
Dy 170 LAKLANAH--VTATCGARNIEFVKSLGDADEVLDYKTPGGAALKSPSGKKYDAVVHCANGI 227

RESULT 27
AAG05067
ID AAG05067 standard; Protein; 267 AA.
XX AAG05067;
AC AAG05067;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 1331.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134258.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.

PR	05-AUG-1999;	99US-0147260.	
PR	06-AUG-1999;	99US-0147303.	
PR	06-AUG-1999;	99US-0147416.	
PR	09-AUG-1999;	99US-0147493.	
PR	09-AUG-1999;	99US-0147935.	
PR	10-AUG-1999;	99US-0148171.	
PR	11-AUG-1999;	99US-0148319.	
PR	12-AUG-1999;	99US-0148341.	
PR	13-AUG-1999;	99US-0148565.	
PR	13-AUG-1999;	99US-0148684.	
PR	16-AUG-1999;	99US-0149369.	
PR	17-AUG-1999;	99US-0149175.	
PR	18-AUG-1999;	99US-0149426.	
PR	20-AUG-1999;	99US-0149722.	
PR	20-AUG-1999;	99US-0149723.	
PR	20-AUG-1999;	99US-0149929.	
PR	23-AUG-1999;	99US-0149902.	
PR	23-AUG-1999;	99US-0149930.	
PR	25-AUG-1999;	99US-0150566.	
PR	26-AUG-1999;	99US-0150884.	
PR	27-AUG-1999;	99US-0151065.	
PR	27-AUG-1999;	99US-0151066.	
PR	27-AUG-1999;	99US-0151080.	
PR	30-AUG-1999;	99US-0151303.	
PR	31-AUG-1999;	99US-0151438.	
PR	01-SEP-1999;	99US-0151930.	
PR	07-SEP-1999;	99US-0152363.	
PR	10-SEP-1999;	99US-0153070.	
PR	13-SEP-1999;	99US-0153758.	
PR	15-SEP-1999;	99US-0154018.	
PR	16-SEP-1999;	99US-0154039.	
PR	20-SEP-1999;	99US-0154779.	
PR	22-SEP-1999;	99US-0155139.	
PR	23-SEP-1999;	99US-0155486.	
PR	24-SEP-1999;	99US-0155659.	
PR	28-SEP-1999;	99US-0156458.	
PR	29-SEP-1999;	99US-0156596.	
PR	04-OCT-1999;	99US-0157117.	
PR	05-OCT-1999;	99US-0157753.	
PR	06-OCT-1999;	99US-0157865.	
PR	07-OCT-1999;	99US-0158029.	
PR	08-OCT-1999;	99US-0158232.	
PR	12-OCT-1999;	99US-0158339.	
PR	13-OCT-1999;	99US-0159293.	
PR	13-OCT-1999;	99US-0159294.	
PR	13-OCT-1999;	99US-0159295.	
PR	14-OCT-1999;	99US-0159329.	
PR	14-OCT-1999;	99US-0159330.	
PR	14-OCT-1999;	99US-0159331.	
PR	14-OCT-1999;	99US-0159637.	
PR	14-OCT-1999;	99US-0159638.	
PR	18-OCT-1999;	99US-0159584.	
PR	21-OCT-1999;	99US-0160741.	
PR	21-OCT-1999;	99US-0160767.	
PR	21-OCT-1999;	99US-0160768.	
PR	21-OCT-1999;	99US-0160770.	
PR	21-OCT-1999;	99US-0160814.	
PR	21-OCT-1999;	99US-0160815.	
PR	22-OCT-1999;	99US-0160980.	
PR	22-OCT-1999;	99US-0160981.	
PR	22-OCT-1999;	99US-0160989.	
PR	25-OCT-1999;	99US-0161404.	
PR	25-OCT-1999;	99US-0161405.	
PR	25-OCT-1999;	99US-0161406.	
PR	26-OCT-1999;	99US-0161359.	
PR	26-OCT-1999;	99US-0161360.	
PR	26-OCT-1999;	99US-0161361.	
PR	28-OCT-1999;	99US-0161920.	
PR	28-OCT-1999;	99US-0161922.	
PR	28-OCT-1999;	99US-0161993.	
PR	29-OCT-1999;	99US-0162142.	
Query Match			8.4%; Score 162; DB 21; Length 267;
Best Local Similarity 26.2%; Pred. No. 2.5e-06;			
Matches 63; Conservative 35; Mismatches 110; Indels 32; Gaps 9;			
Qy	14	GDXAVVKTDSVPDELKEGTALVKVEAVAGNPTDKHIAKIGP-----EGSILGCDIAGT 68	
Db	17	GAAGLEHVQVPVPTPKSNEVCLKLEATSLNPVDWKIQKGMIRPLPRKFCIPATDVAGE 76	
Qy	69	VVKLGPNASTDLKVGDTGFGFVHGASQTDPKNGCAFAEYARVYPLFYKSNLTHSTADEIS 128	
Db	77	VVOVGGGVK-NFRAGDKVVAULSHLG-----GGGLAEFA-----VATEKLTVKRPOEVG 124	
Qy	129	EGPVKNFESAASLPVSLTTAGVSLCHHLGSKMEWHPSTPQHTPLLIWGATAVGQOLIQ 188	
Db	125	AA-----EAAALPVAGLTALQALTNPAGLKLD--GTGKAN-ILVTAASGGVGHYAVQ 174	
Qy	189	VAKHINAYKIVTVASKKHEKLLKSYGADDVFDY----HDAGVIEQIKSKYPNIQHVDAV 245	
Db	175	LAKLANAH--VTATCGARNIEFVKSLGAEVLDYKTPGGAALKSPSGKKYDAVVHCANGI 232	
RESULT 28			
AAG05066			
ID	AAG05066 standard; Protein; 292 AA.		
XX	AAG05066;		
AC	AAG05066;		
XX	AAG05066;		
DT	17-OCT-2000 (first entry)		
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 1330.		
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 1330.		
XX	Protein identification; signal transduction pathway; metabolic pathway;		
KW	hybridisation assay; genetic mapping; gene expression control; promoter;		
KW	termination sequence.		
XX	Arabidopsis thaliana.		
OS	Arabidopsis thaliana.		
XX	EP1033405-A2.		
PN	EP1033405-A2.		
XX	06-SEP-2000.		
PD	06-SEP-2000.		
XX	25-FEB-2000; 2000EP-0301439.		
PF	25-FEB-1999; 99US-0121825.		
XX	25-FEB-1999; 99US-0123180.		
PR	05-MAR-1999; 99US-0123180.		
PR	03-MAR-1999; 99US-0123548.		
PR	23-MAR-1999; 99US-0125788.		
PR	25-MAR-1999; 99US-0125788.		
PR	29-MAR-1999; 99US-0126785.		
PR	01-APR-1999; 99US-0127462.		
PR	08-APR-1999; 99US-0128234.		
PR	16-APR-1999; 99US-0129845.		
PR	19-APR-1999; 99US-0130077.		
PR	21-APR-1999; 99US-0130449.		
PR	23-APR-1999; 99US-0130510.		
PR	28-APR-1999; 99US-0130891.		
PR	30-APR-1999; 99US-0131449.		
PR	30-APR-1999; 99US-0132048.		
PR	04-MAY-1999; 99US-0132407.		
PR	05-MAY-1999; 99US-0132484.		
PR	06-MAY-1999; 99US-0132485.		
PR	06-MAY-1999; 99US-0132486.		
PR	06-MAY-1999; 99US-0132487.		
PR	07-MAY-1999; 99US-0132863.		
PR	11-MAY-1999; 99US-0134256.		
PR	14-MAY-1999; 99US-0134218.		
PR	14-MAY-1999; 99US-0134219.		
PR	14-MAY-1999; 99US-0134221.		
PR	14-MAY-1999; 99US-0134370.		
PR	18-MAY-1999; 99US-0134768.		
PR	19-MAY-1999; 99US-0134941.		
PR	20-MAY-1999; 99US-0135124.		
PR	21-MAY-1999; 99US-0135353.		

PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140351.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147935.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.

PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 8.4%; Score 162; DB 21; Length 292;
 Best Local Similarity 26.2%; Pred. NO. 2.8e-06;
 Matches 63; Conservative 35; Mismatches 110; Indels 32; Gaps 9;

QY 14 GDKAVKTDVSPPELKEGTALVKVEAVAGNPTDMKHTAYKIGP-----EGSILGCDIAGT 68
 DB 42 GAAGLEHVQVVPPTPKSNEVCLEATSLNPDWKIQKGMIRPFLPRKFPICPATDVAGE 101
 QY 69 VVKLGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAYARYVYPLFYKSNLTHSTADEIS 128
 DB 102 VVQVSGVK-NFKAGDKVAVLSHLG-----GGGLAEFA-----VATEKLTVKRPQEVG 149
 QY 129 EGPVKNFESAASLPVSLTTAGVSLCHLHLSKMEWHPSTPQHTPHLLIWGAGATVGOQLIQ 188
 DB 150 AA-----EAAALPVAGLTALQALTNPAGLKLD---GTGKKAN-ILVTAASGGVGHYAVQ 199
 QY 189 VAKHINATKYVTVASKKHEKLLKSYGADDDVFDY---HDAGVIEQIKSKYPNLOHVIDAV 245
 DB 200 LAKLANAH--VTATCGARNIEFVKSLGADDEVLYKTPEGAALKSPGKKYDAVVHCANGI 257

RESULT 29
 AAR97569
 ID AAR97569 standard; Protein; 396 AA.
 AC AAR97569;
 XX
 XX
 DT 13-JAN-1997 (first entry)
 XX
 DE Interleukin-2 receptor associated protein p43.
 XX
 XX p43; interleukin-2 receptor; IL-2; treatment; diagnosis; disease;
 KW signal transduction.
 XX
 XX Homo sapiens.
 XX
 XX WO9621732-A1.
 XX
 XX 18-JUL-1996.
 XX
 XX 23-DEC-1995; 95WO-EP05123.
 XX
 PR 09-JAN-1995; 95EP-0100201.
 XX
 XX (BOEH) BOEHRINGER INGELHEIM INT GMBH.
 XX
 XX Barsoumian EL, Shibuya H, Taniguchi T;
 XX
 XX WPI; 1996-342285/34.
 DR
 DR N-PSDB; AAT31854, AAT31855.
 XX
 XX p43 and related polypeptide(s) which bind to interleukin receptor
 PT - useful for treatment and diagnosis of IL-2 related disease, and to
 PT study mechanism of IL-2 mediated signal transduction
 XX
 XX Claim 7; Page 22-23; 39pp; English.
 XX
 XX A polypeptide having p43-like biological activity, related peptides
 CC and antisense oligonucleotides to the p43 gene are useful in the
 CC treatment and diagnosis of interleukin-2 related diseases. p43 can
 CC also be used to study the mechanism of IL-2 mediated signal
 CC transduction.
 XX
 XX
 SQ Sequence 396 AA;

Query Match 8.3%; Score 158.5; DB 17; Length 396;
 Best Local Similarity 27.8%; Pred. NO. 9.2e-06;
 Matches 71; Conservative 39; Mismatches 96; Indels 49; Gaps 12;

QY 34 LKVEAVAGNPTDWK-----HIAYKIGPEGSI-LQCDIAGTVVVKLGN 75
 DB 75 IVKHAASVNPIDVNMRSYGATANKRDLPHVKIK-GEEFPLTLGRDVGSGVMECGLD 133
 QY 76 ASTDLKVGDTGFGFVHGASQTDPKNGAFAYARYVYPLFYKSNLTHSTADEISGPVK-N 134
 DB 134 VKY-FKPGDEW-----AAVPPMKQGTLSSEFVV-----SGNEVSHKPKSLT 174
 QY 135 FESAASLPVSLTTAGVSLCHLHLSKMEWHPSTPQHTPHLLIWGAGATVGOQLIOVAKHIN 194
 DB 175 HTQAASLPVVALTAW-SAINKVGGLNDKNCCTGKR-----VLILGASGGVGTFAIVMKAWD 229
 QY 195 AYTKITVASKKHEKLLKSYGADDDVFDYHDAGVIEQIKSKYPNLOHVIDAV--SEDSIP 252
 DB 230 AH--VTAVCSQDASELVRKLGADDDVLDYKSGSVEEQLSKLP-FDILDNVGSGSTETWAP 286
 QY 253 EAYKVTADSLPATLL 267
 DB 287 DFLKWSGATYVTLV 301

RESULT 30
 ABB11836
 ID ABB11836 standard; peptide; 413 AA.
 AC ABB11836;
 XX
 XX
 DT 11-JAN-2002 (first entry)
 XX
 DE Human IL-2R-associated protein 43 homologue, SEQ ID NO:2205.
 XX
 XX Human; cytokine; cell proliferation; cell differentiation; growth factor;
 KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
 KW inhibitor; chemotaxis; chemokinesis; thrombolytic; oncogenesis;
 KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
 KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
 KW chronic inflammatory condition; proliferative retinopathy;
 KW atherosclerosis; coronary heart disease; arterial ischaemia;
 KW bone disorder; osteoporosis; vascular growth disorder;
 KW tissue regeneration; wound healing; infection; immune disorder;
 KW cell culture; drug screening; gene therapy; antiinflammatory;
 KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
 KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
 KW antifungal; vulnery; antiulcer.
 XX
 XX Homo sapiens.
 OS
 XX
 XX WO200157188-A2.
 XX
 XX 09-AUG-2001.
 XX
 XX 05-FEB-2001; 2001WO-US03800.
 PF
 XX
 XX 03-FEB-2000; 2000US-0496914.
 PR
 XX 27-APR-2000; 2000US-0560875.
 PR
 XX (HYSE-) HYSEQ INC.
 PA
 XX
 XX Tang YT, Liu C, Drmanac RT;
 PI
 XX WPI; 2001-457740/49.
 XX
 XX N-PSDB; ABA09080.
 DR
 DR
 XX Human proteins and DNA encoding sequences useful for preventing,
 PT treating or ameliorating a medical condition in a mammalian subject
 PT e.g. arthritis and cancer -
 PT
 XX
 XX Claim 20; Page 265; 1963pp; English.
 PS
 XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
 CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
 CC invention also relates to vectors and recombinant host cells comprising a
 CC nucleotide of the invention, methods of producing the novel polypeptides,
 CC

antibodies against the polypeptides, methods of detecting the nucleotides or polypeptides in a sample, and methods of identifying compounds which bind to polypeptides of the invention. Although novel, many of the polypeptides of the invention have homology to known proteins, thereby giving an insight into their probable biological activities, and hence potential therapeutic applications. The polypeptides of the invention may have various activities, including cytokine, cell proliferation or cell differentiation activities; stem cell growth factor activity; haematopoietic regulatory activity; tissue growth factor activity; immunomodulatory activity; activin- or inhibin-related activities; chemotactic or chemokinetic activities; haemostatic, thrombotic or thrombolytic activities; receptor or ligand activities; or may be involved in oncogenesis, cancer cell proliferation or metastasis. Depending on their biological activities, polypeptides and nucleotides of the invention are useful for preventing, treating or ameliorating medical conditions, e.g., by protein or gene therapy. Such conditions include cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell disorders), chronic inflammatory conditions (e.g., asthma or arthritis), proliferative retinopathy, atherosclerosis, coronary heart disease, arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal vascular growth. Polypeptides involved with tissue regeneration and repair (or nucleic acids encoding them) may be used to promote wound healing (e.g., of burns, incisions and ulcers), while those with immunomodulatory activities may be used in the treatment of viral, bacterial and fungal infections in addition to immune disorders. Polypeptides with growth factor activity may be used in cell cultures to promote cell growth. For example, such polypeptides may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease or accidental damage. The polypeptides and nucleotides may also be used in the diagnosis of the above conditions, and in drug screening techniques. The present sequence represents a novel human polypeptide of the invention.

Sequence 413 AA;

Query Match		8.3%; Score 158.5; DB 22; Length 413;
Best Local Similarity		27.8%; Pred. No. 9.8e-06;
Matches		71; Conservative 39; Mismatches 96; Indels 49; Gaps 12;
Qy	34 LKVEAVAGNPTDWK-----	-----HIYKIGPEGSI-LGCDTAGTVVKLGN 75
Db	92 IVKHAASVNPIDVNMBSGYGATALNMKRDPLHVYKIK-GEEFPLTLGRDVSQVVMCEGLD 150	
Qy	76 ASTDLKVGDTGFGFVHGASQTDPKNGAPAEYARVYVPLFYKSNLTHSTADEISSEGVK-N 134	
Db	151 VKY-FKPGDEW-----	AAVPPWQGLSEFVV-----SGNEVSHKPKSLT 191
Qy	135 FESAASLPVSLTTAGVSLCHHLGSKMWHPSPTQHTPLLIWGGATAVGQQLIQVAKHIN 194	
Db	192 HTQAASLPYVALTAW-SAINKVGGLNDKNCCTGKR----	VLILGASGGVGTFAIQVMKAWD 246
Qy	195 AYTQIVTVASKKHEKLLKSYGADDDVFDHAGVIEQIKSKYPNLOHVDAVG---SEDSIP 252	
Db	247 AH--VTAVCSQDASELVKRLGADDVIDYKSGSVEQLKSLKP-PDFILDNVGGSTETWAP 303	
Qy	253 EAYKVTADSLPATLL 267	
Db	304 DFLKNWGGATVTLV 318	

Search completed: January 10, 2004, 10:04:54
Job time : 80 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2004, 09:56:13 ; Search time 37 Seconds
(without alignments)
961.687 Million cell updates/sec

Title: US-10-081-644-2

Perfect score: 1920

Sequence: 1 MSVPTTQKAVIEGKAVK.....ALTEGKGNKKNKYVARL 370

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database :

PIR_76:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1065.5	55.5	376	2 S55149	hypothetical prote
2	1021.5	53.2	376	2 S59418	hypothetical prote
3	1015.5	52.9	368	2 S19414	hypothetical prote
4	264	13.8	313	2 AE1151	oxidoreductase hom
5	260	13.5	313	2 AF1510	oxidoreductase hom
6	259	13.5	317	2 T10824	auxin-induced prot
7	254	13.2	309	2 E86371	zinc-binding oxido
8	218	11.4	353	2 AF3199	quinone oxidoreduc
9	208	10.8	324	2 C83695	quinone oxidoreduc
10	193	10.1	343	1 C70418	probable alcohol d
11	187.5	9.8	322	2 G83766	quinone oxidoreduc
12	187	9.7	334	2 AH3214	zinc-binding oxido
13	176	9.2	331	2 B90457	alcohol dehydrogen
14	175.5	9.1	348	2 T18210	alcohol dehydrogen
15	173.5	9.0	328	2 AG3182	zinc-binding dehyd
16	173	9.0	328	2 E86714	quinone oxidoreduc
17	172.5	9.0	339	2 B83742	alginatase lyase BH0
18	171.5	8.9	338	2 AD2699	zinc-binding dehyd
19	171.5	8.9	338	2 D97481	probable oxidoredu
20	170.5	8.9	336	2 AH3289	NADPH2:quinone red
21	164.5	8.6	329	2 T10203	hypothetical prote
22	164.5	8.6	348	2 F97459	probable quinone o
23	164	8.5	337	2 A82309	quinone oxidoreduc
24	162	8.4	327	2 A12677	quinone oxidoreduc
25	161.5	8.4	328	2 A70871	probable quinone o
26	161	8.4	329	1 FN0448	zeta-crystallin /
27	159.5	8.3	318	2 AC1858	hypothetical prote
28	159	8.3	329	1 CYGP2	zeta-crystallin /
29	158.5	8.3	331	1 A54932	zeta-crystallin /

RESULT 1

S55149

N;Alternate names: hypothetical protein YNL134c - yeast (Saccharomyces cerevisiae)

C;Species: Saccharomyces cerevisiae

C;Date: 08-Jul-1995 #sequence, revision 01-Sep-1995 #text_change 19-Apr-2002

C;Accession: S55149; S59254; S63079

ALIGNMENTS

30	157	8.2	328	2	T21985	hypothetical prote
31	156.5	8.2	8563	2	T30226	polyketide synthas
32	156.5	8.2	10223	2	T30225	polyketide synthas
33	156	8.1	328	2	E70695	probable oxidoredu
34	154	8.0	365	2	S50409	hypothetical prote
35	153.5	8.0	326	2	T36504	probable quinone o
36	153.5	8.0	345	2	S57611	probable NADPH2:qu
37	153.5	8.0	348	1	S32521	alcohol dehydrogen
38	152	7.9	336	2	AH0201	probable zinc-bind
39	151	7.9	325	2	T40264	zinc-binding dehyd
40	151	7.9	325	2	T05166	quinone reductase
41	149.5	7.8	388	2	C75441	probable NADPH qui
42	148.5	7.7	348	2	T49047	quinone reductase-
43	147	7.7	328	2	E83504	probable oxidoredu
44	147	7.7	447	2	T35960	crotonyl-CoA reduc
45	146.5	7.6	338	2	H90403	alcohol dehydrogen
46	146.5	7.6	342	2	S57612	probable NADPH2:qu
47	146.5	7.6	351	2	G86389	probable allyl alc
48	146	7.6	343	2	T51554	quinone oxidoreduc
49	145.5	7.6	350	2	S09475	alcohol dehydrogen
50	144.5	7.5	6260	2	T30228	polyketide synthas
51	143.5	7.5	335	2	AE2174	hypothetical prote
52	143.5	7.5	335	2	E90014	hypothetical prote
53	143.5	7.5	350	1	S52153	alcohol dehydrogen
54	143	7.4	332	2	AE1396	zinc-binding dehyd
55	142.5	7.4	348	2	H71110	probable dehydroge
56	141.5	7.4	337	1	A42654	alcohol dehydrogen
57	141.5	7.4	348	2	G75049	L-threonine 3-dehy
58	141	7.3	341	2	G86037	threonine dehydrog
59	141	7.3	371	2	AB0407	probable zinc-bind
60	140	7.3	341	1	DEECTH	L-threonine 3-dehy
61	140	7.3	379	2	JN0013	synaptic vesicle m
62	139.5	7.3	337	2	G83200	probable oxidoredu
63	139.5	7.3	346	2	T51551	quinone oxidoreduc
64	139	7.2	341	2	P91190	threonine dehydrog
65	138.5	7.2	325	2	G87344	alcohol dehydrogen
66	137.5	7.2	326	2	F95888	probable NADPH2:qu
67	137.5	7.2	349	2	B90285	alcohol dehydrogen
68	137	7.1	359	2	H95892	probable oxidoredu
69	136	7.1	332	2	AH1771	zinc-binding dehyd
70	136	7.1	358	2	T51555	quinone oxidoreduc
71	134.5	7.0	374	1	S17252	alcohol dehydrogen
72	134	7.0	321	2	E87715	quinone oxidoreduc
73	134	7.0	336	2	F86923	probable oxidoredu
74	134	7.0	411	2	H95851	probable alcohol d
75	132	6.9	334	1	S45904	probable NADPH2:qu
76	131.5	6.8	342	2	AD3391	alcohol dehydrogen
77	131.5	6.8	375	1	S54458	alcohol dehydrogen
78	131	6.8	380	2	E84238	quinone oxidoreduc
79	131	6.8	447	2	S72400	trans-2-enoyl-CoA
80	131	6.8	4151	2	G70944	probable polyketid
81	130.5	6.8	348	1	DEBYA2	alcohol dehydrogen
82	130.5	6.8	333	1	A45052	L-iditol 2-dehydro
83	130	6.8	348	1	S32484	L-iditol 2-dehydro
84	129.5	6.7	348	1	S20911	alcohol dehydrogen
85	129	6.7	334	2	E70519	probable oxidoredu
86	128.5	6.7	348	1	DEBYA	alcohol dehydrogen
87	128	6.7	311	2	T51553	quinone oxidoreduc
88	128	6.7	340	2	H95349	alcohol dehydrogen
89	128	6.7	1582	2	E70876	probable polyketid
90	128	6.7	2118	2	S72705	myceroate synth

	Query Match	52.9%;	Score 1015.5;	DB 2;	Length 368;
	Best Local Similarity	53.0%;	Pred. NO. 1.1e-68;		
	Matches 194;	Conservative 60;	Mismatches 109;	Indels 3;	Gaps 2;
Qy	8 KAVITBGDRAVVKTVDSYVPELKEGCTALVKVEAVAGNPDTMKHAYKIGPESGIICDDTA	67			
Dd	2 KAVVIDSGAAVKVEGPVPIPELSEGEFVLIKTAVAGNPDTMWAHIDYIKPGQSILGCDAAG	61			
Qy	68 TVVKLGPNAS--TDLKVGDTGFQFVHGASQTDPKNGAFAEYARVPPLPYKS--NLTHSTA	124			
Dd	62 OIVKGLPAVDPKDFSIGDYIYVFIFIGSSVRFPNSNGAFAEYSASTWYAKSINELKFLCE	121			

```
Qy 125 DEISEGPKVNFESASLPVSLTTAGVSLCHHLGSKMEWHPSTPQHTPLLIWGGATAVGQ 184
Db 122 DVLPAQVRSLEGAATIPVSLTTAGLVTLNGLNLKWEPSTPQNRGPIILLWGGATAVGQ 181
Qy 185 OLIOVAKHINAYTKIVTVASKKHEKLKSYGADDDVFDYHDAGVIEQIKSKYPNLOHVDA 244
Db 182 SLLOLANKNGFTKIIIVASKKHEKLKSYGADQQLFDYHDVIEQIKHKNYINSLVDC 241
Qy 245 VGSDDSIPEAYKVTVADSLPATLLEVPMTIESIEEIRKDNVKIDITLLYRASQEIILG 304
Db 242 VANQNTIQQVYKCAADKQDATVLTNLTENVKKNRQNVITDRTLXSIGHEVFPFG 301
Qy 305 ATRPPASPEYHEATVKVFPKFNPHLNGDIHNMNKFVNSGLDVPALTEGIEKGNKNV 364
Db 302 GITPPADPEARRAATEFVKFNPKNISDQIHHPARVVYKGLYDVPRILEDIKIGKNSGE 361
Qy 365 KYVARL 370
Db 362 KLVAVL 367

RESULT 4
AE1151
oxidoreductase homolog lmo0613 [imported] - Listeria monocytogenes (strain EGD-e)
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AE1151
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Feihl, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,
A.; Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AE1151
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-313 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAC98691.1; PID:g16410002; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
C;Gene: lmo0613
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

Query Match 13.8%; Score 264; DB 2; Length 313;
Best Local Similarity 29.0%; Pred. No. 2.7e-12;
Matches 102; Conservative 56; Mismatches 120; Indels 74; Gaps 18;

Qy 8 KAVIIE---GDKAVVKTDSVPPELKEGTALVKVAVAGNPTDWK---HIAYKIGPE-GS 59
Db 2 KAVVIENGGKEELKEVAMPKAGKQNVKEATSNIPDWKLRGILKQMDWEFPI 61
Qy 60 ILGCDIAGTVVKGPNASTDLKVGDTGFGFVHGASQTDPKNGAPAFAYRVPPLFYKSNL 119
Db 62 ILGWDVAGVISEVGEV-TDMKVGDEVF-----ARPETTFGTVAEYTA-V-----DDHL 109
Qy 120 THSTADEISEGPKVNFESASLPVSLTTAGVSLCHHLGSKMEWHPSTPQHTPLLIWGG 179
Db 110 LAPLEGIS-----FEEAASIFLAGITAWQALFDH--AKL-----QKGEKVLHAGA 154
Qy 180 TAVGQQLIOVAKHINAYTKIVTVASKKHEKLKSYGADDDVFDYHDAGVIEQIKSKYPN 239
Db 155 GGVGTGLAIQLAKHAGA--EVITTSARKNHELKSLGADQVDIYKEV---NFKDVLSDID 208
Qy 240 HVIDAVGSEDSIPEAYKVTVADSLPATLLEVPMTIESIEEIRKDNVKIDITLLYRASQ 299
Db 209 VFEDTMGGQIE-TDSYDVLKEG-TGLRVSIV-----GISNEDRAKKNVTATGWLQPNG 261
Qy 299 QEI-----LLGATRP-----ASPEYHEATVKVFKFN 326
Db 262 EQLKELGKLLANKTIKPIVGAT-PFPSEKGVFDAHALSETHAVGKIVISFN 312

oxidoreductase homolog lmo0613 [imported] - Listeria monocytogenes (strain EGD-e)
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AE1151
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Feihl, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,
A.; Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AE1151
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-313 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAC98691.1; PID:g16410002; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
C;Gene: lmo0613
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

Query Match 13.8%; Score 264; DB 2; Length 313;
Best Local Similarity 29.0%; Pred. No. 2.7e-12;
Matches 102; Conservative 56; Mismatches 120; Indels 74; Gaps 18;

Qy 8 KAVIIE---GDKAVVKTDSVPPELKEGTALVKVAVAGNPTDWK---HIAYKIGPE-GS 59
Db 2 KAVVIENGGKEELKEVAMPKAGKQNVKEATSNIPDWKLRGILKQMDWEFPI 61
Qy 60 ILGCDIAGTVVKGPNASTDLKVGDTGFGFVHGASQTDPKNGAPAFAYRVPPLFYKSNL 119
Db 62 ILGWDVAGVISEVGEV-TDMKVGDEVF-----ARPETTFGTVAEYTA-V-----DDHL 109
Qy 120 THSTADEISEGPKVNFESASLPVSLTTAGVSLCHHLGSKMEWHPSTPQHTPLLIWGG 179
Db 110 LAPLEGIS-----FEEAASIFLAGITAWQALFDH--AKL-----QKGEKVLHAGA 154
Qy 180 TAVGQQLIOVAKHINAYTKIVTVASKKHEKLKSYGADDDVFDYHDAGVIEQIKSKYPN 239
Db 155 GGVGTGLAIQLAKHAGA--EVITTSARKNHELKSLGADQVDIYKEV---NFKDVLSDID 208
Qy 240 HVIDAVGSEDSIPEAYKVTVADSLPATLLEVPMTIESIEEIRKDNVKIDITLLYRASQ 298
Db 209 VFEDTMGGQIE-TDSYDVLKEG-TGLRVSIV-----GISNEDRAKKNVTATGWLQPNG 261
Qy 299 QEI-----LLGATRP-----ASPEYHEATVKVFKFN 326
Db 262 EQLKELGKLLANKTIKPIVGAT-PFPSEKGVFDAHALSETHAVGKIVISFN 312
```

RESULT 5

```
AF1510
oxidoreductase homolog lin0622 [imported] - Listeria innocua (strain Clip11262)
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AF1510
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Feihl, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,
A.; Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AF1510
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-313 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC95854.1; PID:g16413062; GSPDB:GN00178
A;Experimental source: strain Clip11262
C;Genetics:
C;Gene: lin0622
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

Query Match 13.5%; Score 260; DB 2; Length 313;
Best Local Similarity 28.8%; Pred. No. 5.4e-12;
Matches 101; Conservative 60; Mismatches 118; Indels 72; Gaps 18;

Qy 8 KAVIIE---GDKAVVKTDSVPPELKEGTALVKVAVAGNPTDWK---HIAYKIGPE-GS 59
Db 2 KAVVIENGGKEELKEVAMPKAGKQNVKEATSNIPDWKLRGILKQMDWEFPI 61
Qy 60 ILGCDIAGTVVKGPNASTDLKVGDTGFGFVHGASQTDPKNGAPAFAYRVPPLFYKSNL 119
Db 62 ILGWDVAGVISEVGEV-TDMKVGDEVF-----ARPETTFGTVAEYTA-V-----DDHL 109
Qy 120 THSTADEISEGPKVNFESASLPVSLTTAGVSLCHHLGSKMEWHPSTPQHTPLLIWGG 179
Db 110 LAPLEGIS-----FEEAASIFLAGITAWQALFDH--AKL-----QKGEKVLHAGA 154
Qy 180 TAVGQQLIOVAKHINAYTKIVTVASKKHEKLKSYGADDDVFDYHDAGVIEQIKSKYPN 239
Db 155 GGVGTGLAIQLAKHAGA--EVITTSARKNHELKSLGADQVDIYKEV---NFKDVLSDID 208
Qy 240 HVIDAVGSEDSIPEAYKVTVADSLPATLLEVPMTIESIEEIRKDNVKIDITLLYRASQ 299
Db 209 VFEDTMGGQIE-TDSYDVLKEG-TGLRVSIVGISNE---DRAKEKNVTAN-GIWLQPNG 262
Qy 300 EI-----LLGATRP-----ASPEYHEATVKVFKFN 326
Db 263 QLKELGKLLANKTIKPIVGAT-PFPSEKGVFDAHALSETHAVGKIVISFN 312

auxin-induced protein (clone MII-3) - mung bean
C;Species: Vigna radiata (mung bean)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Sep-1999
C;Accession: T10824
R;Chen, J.; Wu, D.; Witham, F.H.; Heuser, C.W.; Arteca, R.N.
submitted to the EMBL Data Library, February 1995
A;Description: Molecular cloning and characterization of auxin-regulated genes from mung
A;Reference number: Z17176
A;Accession: T10824
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-317 <CHE>
A;Cross-references: EMBL:U20808; NID:g1184120; PID:g1184121
A;Experimental source: strain Rwilcz cv. Berken; tissue: hypocotyl; clone MII-3
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
```


C;Genetics:
A;Gene: BH0363
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

Query Match 10.8%; Score 208; DB 2; Length 324;
Best Local Similarity 26.4%; Pred. No. 4.6e-08;
Matches 90; Conservative 53; Mismatches 132; Indels 66; Gaps 15;

Qy 8 KAVIIE---GDKAVKTDVSVPELKEGTALVKVEAVAGNPTDVKHIAIKYKIPGSGIL--- 61
Db 2 KAIWVTAFGGPEHMMWEDVSTPVKENEVLKVKVTSVNFAD---IKARYGRKGTLFFI 58

Qy 62 -GCDIAGTVVKGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAYARVYPPFIYKSNLT 120
Db 59 PGLDAAGYVKGREVS-DIQVGORVIAF-----PKSGSVAEY-----VVAABSLV 103

Qy 121 HSTADEISEGPKVNFESAASLPVSLTTAGVSLCHLGSKMWHPTSPQHTHPILLINGAT 180
Db 104 FPIPDEI-----NFRYAAASPI-----VSFLSH-----RLLYNVAQMERGESVLVHAAAG 148

Qy 181 AVGQOLIQAQKHINAYTKIVTVASKKHEKLLKSYGADDDVFDYHDAGVIEQIKSKYPNLOH 240
Db 149 GVGTTALQMAKLLGAGTIGTVGSADKITAAGSGADEVICYEEDEFTKSV-----QE 201

Qy 241 VIDAVGSEDSIPEYAKVTADSLPATLLEVPMPTIESIPEIRKDNVKIDI-----TLLYRA 296
Db 202 MTNGVGVD-----IILDSVSGSVTEKSLQCLARYGLRVHFGNSSGAIGTITKIDLHA 253

Qy 297 SGQEIIL---LGATFPASPPEVHEATVKVFEINPHLANGDI 334
Db 254 SCRSLVGLSGTTR-KCKPHLLKETARHVL---PYLASGKL 290

RESULT 10
C70418
probable alcohol dehydrogenase (EC 1.1.1.-) - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: C70418
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'N.

Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: C70418
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-343 <AQF>
A;Cross-references: GB:AE000736; NID:g2983763; PIDN:AAC07327.1; PID:g2983768; GB:AE00065
A;Experimental source: strain VF5
C;Genetics:
A;Gene: adh1
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C;Keywords: NAD; oxidoreductase
F;26-332/Domain: long-chain alcohol dehydrogenase homology <LADH>

Query Match 10.1%; Score 193; DB 1; Length 343;
Best Local Similarity 28.7%; Pred. No. 6.7e-07;
Matches 77; Conservative 43; Mismatches 94; Indels 54; Gaps 14;

Qy 8 KAVIIEG---DKAVKTDVSVPELKEGTALVKVEAVAGNPTD-WKHI-AVKIGPE-GSI 60
Db 2 KAVILTGFGGIENLKYVEDFPKPEKDEVILIRKVALNHLDIWVRMGALPVKPELPHI 61

Qy 61 LGCDIAGTVVKGPNASTDLKVG-----DTGFGFVHGASQTD-----PKNG 101
Db 62 LGSVDSVGWEKVG-SLVKVNKEGEVIAIPGLSGVCWCQSGRDNHCKDYDILGLKNKG 120

Qy 102 AFAEYARVYPPFIYKSNLTHTADEISEGPKVKN--PESASLPVSLTTAGVSLCHLHLSK 159
Db 121 GIABYAV-----PARNVTKPK-KNLSFEEAASYPITFLT-----VNNALVDK 162

Qy 160 MEMHPSTPQHTHPILLINGGATAVGQQLIQVAKHINAYTKIVTVASKKHEKLLKSYGADDV 219

Db 163 AQIKPYS-----RVFIWAGSSGVGAGIQLAKAFNAFV-ITTAGNEBKAKCKELGADLV 216
Qy 220 FDVHDAGVIEQIKSKY-PNLQHVVIDAVG 246
Db 217 FHHYKDDVVKVREVPKGVGVVDVVDHIG 244

RESULT 11
G83766
quinone oxidoreductase BH0935 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: G83766
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, N.; Masui, F.; Hizen
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: G83766
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-322 <STO>
A;Cross-references: GB:AP001510; GB:BA000004; NID:g10173440; PIDN:BA04654.1; GSPDB:GN001
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH0935
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

Query Match 9.8%; Score 187.5; DB 2; Length 322;
Best Local Similarity 27.0%; Pred. No. 1.6e-06;
Matches 79; Conservative 44; Mismatches 83; Indels 87; Gaps 14;

Qy 8 KAVIIE-GDKAVKVT-DVSVPELKEGTALVKVEAVAGNPTD-----WKHIAY 52
Db 2 KAVVYEQGDPSVLKVDVDPKPTIGTDLVINKASGTNPVDTYPRKGIRQVQFPHPH 61

Qy 53 KIGPEGSLGCDIAGTVVKGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAYARVYPP 112
Db 62 -----FDVAGEIVEIGSDIE-NMQVGDRVWA-----TNIK-GASAEVALIPEH 102

Qy 113 LPYKSNLTHTADEISEGPKVPKESASLPVSLTTAGVSLCHLGSKMWHPTSPQHTHP 172
Db 103 LLFP--LPKSV-----SYEEGALAMTPTAHL--FDRGRLOKGET 141

Qy 173 LLIWGATAVGQQLIQVAKHINAYTKIVTVASKKHEKLLKSYGADDDVFDVHDAGVIEQIK 232
Db 142 VLYGSGVAGHAIQLAKRAGA-TVITTAGDREKGEIAKQAGADQVIFKE----- 192

Qy 233 SKYPNLQHVVIDAVGSEDSIPEYKVTADSLPATLLEVPMPTIE-SIPHEIRKD 284
Db 193 -----ESVVDVAGK-----ATNGGVPLILDMSLSENMAQD 223

RESULT 12
AH3214
zinc-binding oxidoreductase Atu5447 [imported] - Agrobacterium tumefaciens (strain C58; I
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AH3214
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.;
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McEllell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, M.;
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AH3214
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-334 <KUR>
A;Cross-references: GB:AE008687; PIDN:AAL46134.1; PID:g17743901; GSPDB:GN00188
A;Experimental source: strain C58 (Dupont)

Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; MUID:21608550; PMID:11743193

A:Accession: AD2699

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-338 <KUR>

A:Cross-references: GB:AE008688; PIDN:AAL42010.1; PID:gl7739384; GSPDB:GN00186

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: Atu0996

A:Map position: circular chromosome

Query Match 8.9%; Score 171.5; DB 2; Length 338;
Best Local Similarity 28.6%; Pred. No. 2.7e-05;
Matches 64; Conservative 36; Mismatches 89; Indels 35; Gaps 9;

Qy 17 AVVKTDSVPELKEGTALVKVEAGNPTDVKHAYKIGPEGS---ILGCDIAGTVVVKLG 73

Db 17 ALIDIDLAQPAKGDHILVEKAVSNVPDVK-VRRNQSPENGATRVLGFDSAGVVKAVG 75

Qy 74 PNASTDILKVGDTGF--GFVHGASQTDPKNGAFAYARVYPLFYKSNLTHSTADEISEGP 131

Db 76 DRVSL-FKPGDEVFYAGVIN-----RPGSNSEF-----HLVDERIVGAK 113

Qy 132 VK--NFSASALPVSLLTAGVSLCHLGSKNMHPSTPQHTHPLLWGATAVGQQLIQV 189

Db 114 PKSLNFEEAALPLTAYETLFDRLRVK----EPVPGAANAVLVITGGAGVGSIAIQL 169

Qy 190 AKHINATKIVTVASKKHEKLLKSYGADVDYHDAGVIEQIKS 233

Db 170 LRALDTLTVIATASRPETMEWVKELGAHVVD-HGKPIAPQVEA 212

RESULT 19

D97481

Probable oxidoreductase (PA3567) [imported] - Agrobacterium tumefaciens (strain C58, Cent

C:Species: Agrobacterium tumefaciens

C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002

C:Accession: D97481

A:Liiv, F.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,

A:Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum

A:Reference number: A97359; MUID:21608551; PMID:11743194

A:Accession: D97481

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-338 <KUR>

A:Cross-references: GB:AE007869; PIDN:AAK86805.1; PID:gl5156011; GSPDB:GN00169

C:Genetics:

A:Gene: AGR C 1830

A:Map position: circular chromosome

Query Match 8.9%; Score 171.5; DB 2; Length 338;
Best Local Similarity 28.6%; Pred. No. 2.7e-05;
Matches 64; Conservative 36; Mismatches 89; Indels 35; Gaps 9;

Qy 17 AVVKTDSVPELKEGTALVKVEAGNPTDVKHAYKIGPEGS---ILGCDIAGTVVVKLG 73

Db 17 ALIDIDLAQPAKGDHILVEKAVSNVPDVK-VRRNQSPENGATRVLGFDSAGVVKAVG 75

Qy 74 PNASTDILKVGDTGF--GFVHGASQTDPKNGAFAYARVYPLFYKSNLTHSTADEISEGP 131

Db 76 DRVSL-FKPGDEVFYAGVIN-----RPGSNSEF-----HLVDERIVGAK 113

Qy 132 VK--NFSASALPVSLLTAGVSLCHLGSKNMHPSTPQHTHPLLWGATAVGQQLIQV 189

Db 114 PKSLNFEEAALPLTAYETLFDRLRVK----EPVPGAANAVLVITGGAGVGSIAIQL 169

Qy 190 AKHINATKIVTVASKKHEKLLKSYGADVDYHDAGVIEQIKS 233

Db 170 LRALDTLTVIATASRPETMEWVKELGAHVVD-HGKPIAPQVEA 212

RESULT 20

AH3289

NADPH2:quinone reductase (EC 1.6.5.5) [imported] - Brucella melitensis (strain 16M)

C:Species: Brucella melitensis

C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 03-Jun-2002

C:Accession: AH3289

R:DelVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, T.

A:Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis

A:Reference number: AD3252; PMID:11756688

A:Accession: AH3289

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-336 <KUR>

A:Cross-references: GB:AE008917; PIDN:AAL51483.1; PID:gl7982195; GSPDB:GN00190

A:Experimental source: strain 16M

C:Genetics:

A:Gene: BMEI0302

A:Map position: 1

C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

C:Keywords: oxidoreductase

Query Match 8.9%; Score 170.5; DB 2; Length 336;
Best Local Similarity 27.8%; Pred. No. 3.2e-05;
Matches 68; Conservative 33; Mismatches 101; Indels 43; Gaps 11;

Qy 24 SVPELKEGTALVKVEAGNPTDVKHAYKIG-----PEGS--ILGCDIAGTVVVKLG 77

Db 31 AVPEKPEGEILVRVRAAGVNRPD---VLQROQGYAPPGPGASDIPGLEIAGDIVALGHGVK 87

Qy 78 TDLKVGDTGTFVHGASQTDPKNGAFAYARVYPLFYKSNLTHSTADEISEGPVKNFES 137

Db 88 R-FRKGDDQVALLAG-----GGYAEYAVVH-----ESNALPLPSG--FGYIE 126

Qy 138 AASLPVSLTTAGVSLCHLGSKNMHPSTPQHTHPLLWGATAVGQQLIQVAKHINAYT 197

Db 127 AAAPETFTFVHNVFERGLK-----EGEVELVHGSSGIGTTAIQAKAFGA-T 176

Qy 198 KIVTVASKKHEKLLKSYGADVDYHDAGVIEQIKSKY--PNLQHVIVDAGSEDSIPEAY 255

Db 177 VITTAGSKKCDACVKLGADRAINHYEEDFVAVVKEMTGKGVVDVILDMVGG-DYVERN 235

Qy 256 KVTAD 260

Db 236 KAAAE 240

RESULT 21

T10203

hypothetical protein F25G13.100 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Oct-1999

C:Accession: T10203

R:Bevan, M.; Pohl, T.; Weizensegger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Lemcke,

A:submitted to the Protein Sequence Database, June 1999

A:Reference number: Z16991

A:Accession: T10203

A:Molecule type: DNA

A:Residues: 1-329 <BEV>

A:Cross-references: EMBL:AL079349; GSPDB:GN00062; ATSP:F25G13.100

A:Experimental source: cultivar Columbia; BAC clone F25G13

C:Genetics:

A:Gene: ATSP.F25G13.100

A:Map position: 4

A:Introns: 22/3; 70/1; 100/3

C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; MUID:21608550; PMID:11743193

A:Accession: A12677

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-327 <KUR>

A:Cross-references: GB:AB008688; PIDN:AAL41839.1; PID:g17739198; GSPDB:GN00186

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: qor

A:Map position: circular chromosome

C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

Query Match 8.4%; Score 162; DB 2; Length 327;
Best Local Similarity 24.0%; Pred. No. 0.00013;
Matches 86; Conservative 57; Mismatches 131; Indels 84; Gaps 17;

QY 14 GDKAVKTVDSVPBELKEGTALVKVEAVAGNPTDWMKHIAYKIG-----PEGS--ILGCDIAG 67
DB 11 GPEVMQSKAPLPKFPARGELILVKVEAGVNPDP---VAQRQGIYPPFKGASPLIGLEIAG 67
QY 68 TVVKLGNASTDLKVGDTGFGFVHGASQTDPKNGAFAYARVYPPFLFYKSNLTHSTADEI 127
DB 68 EEWALGEGVD-EFKLGDVKVCALANG-----GGYAEYCAV-----PAGQA 105
QY 128 SEGVPKMFES--AASLPVSLTTAGVSLCHHLGSKMWHPTPQHTHPLLIWGATAVGQQ 195
DB 106 LPFPF-KGYDVAKAAALPETFTFWANLFWQWAG-----LTEGET--VLHGGTSGIGTT 155
QY 186 LIQVAKHINAYTKIVTVASKKHEKLKSYGADDDVFDYHDAVGIQIKSKY--PNLQHV 243
DB 156 AIQLAKAFGEAVYATAGSAKCEACVK-LGTYKRAINREDFEIVKSETGCGKGVVDVLD 214
QY 244 AVGSEDSIPKAYKVTADSLPATLLEVVPMTIESIPEIRKDNVKIDITLYRASGQSI-- 301
DB 215 MIGA-----AY-----FEKLAALAKDGCLSLIAFLGGATAEKVDL 250
QY 302 -----LLGATFPASPPEYHEA-TVKPKVFINPLNNGDIIHMKIVFNSGLDDV 349
DB 251 RPIMVKRLTVGTSMRPTADEKRAIRDELVEQWPLIESGKVPVINVRFVT--LEEV 306

RESULT 25

A70871

Probable quinone oxidoreductase - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C:Accession: A70871

R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: A70871

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-328 <COL>

A:Cross-references: GB:AL021184; GB:AL123456; PIDN:CAAL5984.1; PID:g279138

A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: qor

C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

F;25-318/Domain: long-chain alcohol dehydrogenase homology <LADH>

Query Match 8.4%; Score 161.5; DB 2; Length 328;
Best Local Similarity 29.5%; Pred. No. 0.00015;
Matches 72; Conservative 36; Mismatches 95; Indels 41; Gaps 13;

QY 14 GDKAVK-TDVSVPBELKEGTALVKVEAVAGNPTDWMKHIAYKIGPE-GSILGCDIAGTVVK 71

DB 10 GGPGLRHRVDQPOPOPGHGGELLIKABAIGVNFIDTIFRSQYPRELPFVIGSEVCGTVEA 69
QY 72 LGPN---ASTDLKVGDTGFGFVHGASQTDPKNGAFAYARVYPPFLFYKSNLTHSTADEIS 128
DB 70 VGPVTAADTAISVGDR---VVSAS---ANGAYAEFCTA-----PASLTAKVPDDVT 115
QY 129 EGPVKNFESAASLPVSLTTAGVSLCHHLGSKMWHPTPQHTHPLLIWGATAVGQOLIQ 188
DB 116 S-----EVAASALLKGLTA-----HYLLKSV--YPVKRGDT--VLVHAGAGGVGLITQ 160
QY 189 VAKHINAYTKIVTVASKKHEKLKSYGADDDVFDYH-----AGVIEQIKSKYPNLQHV 243
DB 161 WATHLGRV-ITTVSTAERAKLSKDAGADVLDYPEDAWQFAGRVRELTTGG-TGVOAVVD 218
QY 244 AVGS 247
DB 219 GVGA 222

RESULT 26

PN0448

zeta-crystallin / quinone reductase (NADPH) (EC 1.6.-.-) - human

C:Species: Homo sapiens (man)

C:Date: 31-Dec-1993 #sequence_revision 07-Jul-1995 #text_change 11-Jun-1999

C:Accession: PN0448; A54672

R:Gonzalez, P.; Rao, P.V.; Zigler Jr., J.S.

Biochem. Biophys. Res. Commun. 191, 902-907, 1993

A:Title: Molecular cloning and sequencing of zeta-crystallin/quinone reductase cDNA from

A:Reference number: PN0448; MUID:93221534; PMID:8466529

A:Accession: PN0448

A:Molecule type: mRNA

A:Residues: 1-329 <CON>

A:Cross-references: GB:S58039; NID:9299369; PIDN:AAB26039.1; PID:g299370

A:Experimental source: liver

A:Note: Translation of initiator Met is not shown; the authors translated the codon ATC 1

R:Gonzalez, P.; Rao, P.V.; Zigler Jr., J.S.

Genomics 21, 317-324, 1994

A:Title: Organization of the human zeta-crystallin/quinone reductase gene (CRY2).

A:Reference number: A54672; MUID:94375054; PMID:8088825

A:Accession: A54672

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-37 <GO2>

A:Cross-references: GB:L31521

C:Comment: This protein is present at low (enzymatic) levels in this species, in contrast

C:Genetics:

A:Gene: GDB:CRY2

A:Cross-references: GDB:139194; OMIM:123691

A:Map position: lp31-lp22

C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

C:Keywords: eye lens; NAD; oxidoreductase

F;33-318/Domain: long-chain alcohol dehydrogenase homology <LADH>

F;152-181/Region: beta-alpha-beta NADP nucleotide-binding fold

Query Match 8.4%; Score 161; DB 1; Length 329;

Best Local Similarity 24.8%; Pred. No. 0.00016;

Matches 73; Conservative 50; Mismatches 95; Indels 76; Gaps 14;

QY 19 VKTDSVPBELKEGTALVKVEAVAGNPTDWMKHIAYKIGPEGSIL-----GCDIAGTVVKLGP 74

DB 24 LRSADIAPIPKDHQVLKLVHACGVNPE-TYIRSGYTSRKPLLPYTPGSDVAGVIEAVGD 82

QY 75 NASTDLKVGDTGFGFVHGASQTDPKNGAFAYARVYPPFLFYKSNLTHSTADEISEGPVK- 133

DB 83 NASA-FKGDVFE-----TSSTISGGYAEYALAADHTVYKL-----PEKL 121

QY 134 NFESAASLPVSLTTAGVSLCHHLGSKMWHPTPQHTHPLLIWGATAVGQOLIQVAKHI 193

DB 122 DFKQGAATGIPYTAIALIHSAVKA-----GESVLVHGASGGVGLAACQIAR-- 170

QY 194 NAY-TKIV-TVASKKHEKLKSYGADDDVFDYHDAVGIQIKSKYPNLQHVDAVGSSEI 251

Db 171 -AYGLKILGTAGTEGQKVLQNGAHEVFNHREVNYYIDKIK-KY-----VGEKG-- 217
Qy 252 PEAYKVTADSLPATLLEVVPMTIESPEERKONKVIDITLLYRASGOEILLGA 305
Db 218 -----IDIIEMLANVLSKDLULSH-GGRVIVGS 248

RESULT 27
AC1858
hypothetical protein all0412 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AC1858
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. strain PCC 7120
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AC1858
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-318 <CUR>
A;Cross-references: GB:BA000019; PID:BA072370.1; PID:gl7129757; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

Query Match 8.3%; Score 159.5; DB 2; Length 318;
Best Local Similarity 30.0%; Pred. No. 0.0002;
Matches 63; Conservative 25; Mismatches 87; Indels 35; Gaps 7;
Qy 20 KTDVSVPELKEGTALVKVEAVAGNPTDKHIAKIGPE---GSLGCDIAGTVVKLGPN 76
Db 17 EVENKPTPANNEVLVKVYATSNPACDGMRRGFFGRVQLPALGLDVGSGVVEAGENV 76
Qy 77 STDCLKVGTGFGFVHGASQTDPKNGAPAEYARVVPPLFYK--SNLTHSTADEISEGVPKN 134
Db 77 -RDFQVGDE---VYVAIPHELGGANAEYHAPSSMAKPPNNISHL----- 119
Qy 135 FESAASLPVSLTAGVSLCHLGSKMEWHPSPTQHTPLLIWGGATAVGQQLIQVAKHI 194
Db 120 --BAATVPVAGGTAAWALITR-----ANLQVGETVLHGGAGGVGTFAIQAKAG 168
Qy 195 AYTKIVTVASKKHEKLLKSYGADDVPDYHD 224
Db 169 AY--VYTCGGYDIDFVKSGADRAIDYRN 196

RESULT 28
CYGPZ
zeta-crystallin / quinone reductase (NADPH) (EC 1.6.-.-) - guinea pig
N;Alternate names: quinone reductase (NADP)
C;Species: Cavia porcellus (guinea pig)
C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 11-Jun-1999
C;Accession: J50230
R;Rodokanaki, A.; Holmes, R.K.; Borraas, T.
Gene 78, 215-224, 1989
A;Title: Zeta-crystallin, a novel protein from the guinea pig lens is related to alcohol dehydrogenase
A;Reference number: J50230; MUID:89378748; PMID:2777081
A;Accession: J50230
A;Molecule type: mRNA
A;Residues: 1-329 <ROD>
A;Cross-references: GB:M26936; NID:gl91252; PIDN:AAA37035.1; PID:g305333
A;Experimental source: strain 13/N
A;Note: the sequences of seven fragments (79-98, 95-106, 106-114, 209-215, 264-275, 291-300) of zeta-crystallin are identical to those of alcohol dehydrogenase
C;Comment: This protein is distantly related to mammalian and yeast alcohol dehydrogenases
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C;Keywords: eye lens; NADP; oxidoreductase
F;152-181/Region: beta-alpha-beta NADP nucleotide-binding fold

Query Match 8.3%; Score 159; DB 1; Length 329;
Best Local Similarity 24.2%; Pred. No. 0.00023;
Matches 71; Conservative 50; Mismatches 96; Indels 76; Gaps 14;
Qy 19 VKTDSVPPELKEGTALVKVEAVAGNPTDKHIAKIGPEGSIL-----GCDIAGTVVKLGPN 74
Db 24 VQSDVAVPIPKDHQVLKIVHACINPVE-TYIRSGTYTRIPLLPYTPGTDVAGVBSIGN 82
Qy 75 NASTDLKVGDTGFGFVHGASQTDPKNGAPAEYARVVPPLFYKSNLTHSTADEISEGVPK- 133
Db 83 DVSA-FKKGRVP-----TTSTISGGYAEYALASDHTVYRL-----PEKL 121
Qy 134 NFESAASLPVSLTAGVSLCHLGSKMEWHPSPTQHTPLLIWGGATAVGQQLIQVAKHI 193
Db 122 DFRQGAAGIPYPTA---CRAL-----FHSARAKAGESVLVHGASGVGLAAACQIAR-- 170
Qy 194 NAY-TKIV-TVASKKHEKLLKSYGADDVPDYHDAGVIEQIKSKYPNLQHVIVDAVGSEDSI 251
Db 171 -AYGLKVLGTAGTEGQKVLQNGAHEVFNHRAHYIDEIKK-----SIGEKG-- 217
Qy 252 PEAYKVTADSLPATLLEVVPMTIESPEERKONKVIDITLLYRASGOEILLG 304
Db 218 -----VDVIIEMLANVLSNLDKLL-SCGRVIVG 247

RESULT 29
A54932
zeta-crystallin / quinone reductase (NADPH) (EC 1.6.-.-) - mouse
C;Species: Mus musculus (house mouse)
C;Date: 25-Apr-1995 #sequence_revision 07-Jul-1995 #text_change 11-Jun-1999
C;Accession: A54932; S42273
R;Gonzalez, P.; Hernandez-Calzadilla, C.; Rao, P.V.; Rodriguez, I.R.; Ziegler Jr., J.S.; I.
Mol. Biol. Evol. 11, 305-315, 1994
A;Title: Comparative analysis of the zeta-crystallin/quinone reductase gene in guinea pig
A;Reference number: A54932; MUID:94224126; PMID:8170370
A;Accession: A54932
A;Molecule type: mRNA
A;Residues: 1-331 <CON>
A;Cross-references: GB:S70056; NID:g546493; PIDN:AAB30620.1; PID:g546494
A;Experimental source: liver
A;Note: sequence extracted from NCBI backbone (NCBIN:147625, NCBIPI:147626)
A;Note: translation of initiator Met is not shown
R;Joernvall, H.; Persson, B.; du Boie, G.C.; Laverre, G.C.; Chen, J.H.; Gonzalez, P.; Rao, P.
FEBS Lett. 322, 240-244, 1993
A;Title: zeta-Crystallin versus other members of the alcohol dehydrogenase super-family.
A;Reference number: S42272; MUID:93252077; PMID:8486156
A;Contents: annotation
C;Comment: This protein is a major soluble protein of the lens in this species and is ex
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C;Keywords: eye lens; NADP; oxidoreductase
F;33-320/Domain: long-chain alcohol dehydrogenase homology <LADH>
F;152-181/Region: beta-alpha-beta NADP nucleotide-binding fold

Query Match 8.3%; Score 158.5; DB 1; Length 331;
Best Local Similarity 22.4%; Pred. No. 0.00025;
Matches 72; Conservative 53; Mismatches 119; Indels 77; Gaps 13;
Qy 19 VKTDSVPPELKEGTALVKVEAVAGNPTDM--KHIAKIGPEGSIL-GCDIAGTVVKLGPN 75
Db 24 LQSDVVVPVQSHQVLKIVHACGNVPVETVIRSGYSRKPALPYTPGSDVAGIESVGDK 83
Qy 76 ASTDLKVGDTGFGFVHGASQTDPKNGAPAEYA-----RVYPPPLFYKSNLTHSTADEISEG 130
Db 84 VSA-FKKGRVFCY-----STVSGGYAEFALAADTIYF-----LPE- 119
Qy 131 PVKNFESAASLPVSLTAGVSLCHLGSKMEWHPSPTQHTPLLIWGGATAVGQQLIQVA 190
Db 120 -TLNFRQGAALGIPYPTA---CRAL-----FHSARAKAGESVLVHGASGVGLATQIA 169
Qy 191 KHINAYTKIVTVASKKHEKLLKSYGADDVPDYHDAGVIEQIKSKYPNLQHVIVDAVGSEDS 250
Db 170 R-AHGLKVLGTAGSEGGKVLQNGAHEVFNHKEANYIDIKM-----SVGDKDK 218

QY 251 IPBAYKVTADSLPATLLEVPMTIESPEIRKDNVKNIDITLLYRASGOEILLGATRPFA 310
Db 219 -----GVDVLEEMANENSLDKLL--SHGRRVVVVGCRGPI 254
QY 311 SPEYHEATVKVVFPHPLANN 331
Db 255 EINPRDTMAKETSIIIGVSLSS 275

RESULT 30

T21985
hypothetical protein F39B2.3 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 31-Jan-2000
C/Accession: T21985
R/Dobson, R.
submitted to the EMBL Data Library, March 1997
A/Reference number: Z19498
A/Accession: T21985
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-328 <WIL>
A/Cross-references: EMBL:Z92834; PIDN:CAB07384.1; GSPDB:GN00019; CESP:F39B2.3
A/Experimental source: clone F39B2
C/Genetics:
A/Gene: CESP:F39B2.3
A/Map position: 1
A/Introns: 118/1; 145/3
C/Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

Query Match 8.2%; Score 157; DB 2; Length 328;
Best Local Similarity 26.7%; Pred. No. 0.00032;
Matches 67; Conservative 44; Mismatches 90; Indels 50; Gaps 13;
QY 5 TTQKAVIIE--GDKAVYK--TDVSVPELKEGTALVKVEAVAGNPTDWMKHAYKIGPEGS-- 59
Db 8 SSMRAAVVRRFGAPDVEIAVESDMPLEKQVLVRNYAAGVNPVD---TYIRAGQYKLP 64
QY 60 ----ILGCDIAGTVVKLGPNASTDLKVGDTGFGFVHGASQTDPKNGAPAFAYARVYPLFY 115
Db 65 NLPVPGKGAGFVELVGSVK-NVKVGDR----VMYVGEAD----STAEYVAVNRP-- 113
QY 116 KSNLTHSTADIEGPGPVKNFESAASLPVSLTTAGVSLCHLGSKMWHPSTPQHTPLLI 175
Db 114 -----ELPEG--VSFEAGSLGVPYLTAYRALFHLAGAKT-----GDVILV 152
QY 176 WGGATAVGOQLIQVA--KHINAYTKIVTVASKKHEKLKSYGADDDVFDYHDAGVIEQIKS 233
Db 153 HGASGGVGSALMQLAARNIEA---VGTAGSADGIRFVKSLGARNVYNHSDKQYVSKMKN 209
QY 234 KYP-NLQHVIV 243
Db 210 DYPGFGNHIFE 220

Search completed: January 10, 2004, 10:08:01
Job time : 41 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2004, 08:41:33 ; Search time 23 Seconds
(without alignments)
756.517 Million cell updates/sec

Title: US-10-081-644-2

Perfect score: 1920

Sequence: 1 MSVPTQKAVIEGDKAVK.....ALTEGIKEGKNVYVARL 370

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1065.5	55.5	376	1 YNN4 YEAST	P39122 saccharomyc
2	1021.5	53.2	376	1 YL60 YEAST	P54007 saccharomyc
3	1015.5	52.9	368	1 YC22 YEAST	P25608 saccharomyc
4	370.5	19.3	297	1 TOXD COCCA	P54006 cochlidiobol
5	254	13.2	309	1 QORL ARATH	O92uc1 arabidopsis
6	175.5	9.1	348	1 ADH2 CANAL	O94038 candida alb
7	161	8.4	329	1 QOR HUMAN	Q86257 homo sapien
8	159	8.3	329	1 QOR CAVPO	P14115 cavia porce
9	158.5	8.3	331	1 QOR MOUSE	P47199 mus musculu
10	155	8.1	340	1 QOR LEIAM	P28665 leishmania
11	154	8.0	365	1 YM27 YEAST	P28625 saccharomyc
12	153.5	8.0	345	1 P1 ARATH	Q39172 arabidopsis
13	153.5	8.0	348	1 ADH1 KLUMA	Q87288 kluyveromyc
14	152.5	7.9	348	1 ADH2 PICST	O33309 pichia stip
15	149.5	7.8	348	1 ADH1 PICST	O00097 pichia stip
16	148	7.7	330	1 QOR BOVIN	O97764 bos taurus
17	148	7.7	330	1 QOR LANGU	Q28452 lama guanac
18	146	7.6	343	1 P2 ARATH	Q39173 arabidopsis
19	145.5	7.6	350	1 ADH1 KLUMA	P20369 kluyveromyc
20	143.5	7.5	350	1 ADH1 CANAL	P43067 candida alb
21	142.5	7.4	348	1 TDH PYRHO	O58389 pyrococcus
22	142	7.4	341	1 TDH ECOL6	O8fca2 escherichia
23	141.5	7.4	337	1 ADH1 BACST	P12311 bacillus st
24	141.5	7.4	348	1 TDH PYRAB	Q9uyx0 pyrococcus
25	141	7.3	341	1 TDH ECOS7	Q8xej1 escherichia
26	141	7.3	341	1 TDH SHIFL	P59409 shigella fl
27	140.5	7.3	347	1 ADH2 KLUMA	Q9p4c2 kluyveromyc
28	140	7.3	341	1 TDH ECOLI	P07913 escherichia
29	140	7.3	341	1 TDH SHEON	Q8e8j1 shewanella
30	140	7.3	379	1 VATI TORCA	P19333 torpedo cal
31	137.5	7.2	348	1 TDH PYRPU	Q8u259 pyrococcus
32	134.5	7.0	334	1 ADH3 KLUMA	P49384 kluyveromyc
33	132	6.9	334	1 QOR YEAST	P38230 saccharomyc

ALIGNMENTS

RESULT 1
YNN4_YEAST
ID YNN4_YEAST STANDARD; PRT; 376 AA.
AC P53912;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 41.2 kDa protein in PPRI-TOM22 intergenic region.
GN YNL134C OR N1214 OR N1847.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

P07246 saccharomyc
Q06004 bacillus su
Q02912 bombyx mori
P49383 kluyveromyc
Q31186 rhizobium m
P42328 bacillus st
Q82214 salmonella
Q82152 salmonella
P00331 saccharomyc
P00330 saccharomyc
P30359 nicotiana t
Q9K162 vibrio chol
O82515 medicago sa
P27867 rattus norv
P42327 bacillus st
P59410 vibrio para
Q38707 apium grave
P38113 saccharomyc
Q99536 homo sapien
P54202 emericella
P39451 escherichia
P50381 sulfolobus
P30360 nicotiana t
Q64442 mus musculu
Q82jnz2 yersinia pe
P31656 medicago sa
Q31776 bacillus su
P00332 schizosacch
Q91233 streptomyces
Q52998 rhizobium m
P39462 sulfolobus
Q45697 caenorhabdi
P42734 arabidopsis
P80715 streptococc
P80715 streptococc
Q00736 homo sapien
P38105 escherichia
Q49482 arabidopsis
Q57517 haemophilus
P31657 populus del
Q04894 saccharomyc
P19212 neurospora
Q92rri1 fragaria an
P35630 entamoeba h
Q03132 saccharopol
P39450 pasteurella
P07846 ovis aries
Q8d442 vibrio vuln
P72334 rhodobacter
P40783 salmonella
Q98317 rhizobium l
Q9K714 bacillus ha
Q02972 arabidopsis
P25437 escherichia
Q43137 styloeanthe
P25377 saccharomyc
Q10094 schizosacch
P42754 petroselinu

```
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C; PubMed=8619318;
RX MEDLINE=96109932; Bussereau F., Jacquet M.;
RA Mallet L., Bussereau F., Jacquet M.;
RT "A 43.5 kb segment of yeast chromosome XIV, which contains MFA2,
RP ME2, CAP/SRV2, NAM9, FKBI/FP1/RBP1, MOM22 and CPT1, predicts an
RT adenosine deaminase gene and 14 new open reading frames.";
RL Yeast 11:1195-1209(1995).
CC -!- SIMILARITY: BELONGS TO THE YCR102C/YLR460C/YNL134C FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z46843; CAA86891.1; -.
CC EMBL; Z71410; CAA96016.1; -.
CC PIR; S55149; S55149.
CC SGD; S0005078; YNL134C.
CC InterPro; IPR002085; Adh_zn_family.
CC Pfam; PF00107; ADH_zinc_N; 1.
CC Hypothetical protein.
CC SEQUENCE 376 AA; 41164 MW; AE39BBBCDA46C3B99 CRC64;
CC -----
Query Match 55.5%; Score 1065.5; DB 1; Length 376;
Best Local Similarity 54.0%; Pred. No. 1.4e-69;
Matches 201; Conservative 59; Mismatches 109; Indels 3; Gaps 2;
Qy 2 SVPTQKAVIIEGDKAVVKTDSVPELKEGTALVKVEAVAGNPTDWHKIAVKGIGESIL 61
Db 4 SIPETMKAVIENGKAVVQDIPPELEEGFVLKTVAVAGNPTDWHKIDFKIGPGCALL 63
Qy 62 GCDIAGTVVVKLGPNAS--TDLKVGDGTGFGFVHGASQTDPKNGAFAYARVYPPFYK--SN 118
Db 64 GCDAAQGVVVKLGPNVDAARFAIGDIYGVHIGASVRFPSNGAFAYSAISSETAYKPARE 123
Qy 119 LTHSTADSEISGPKVNFESAASLPVSLTTAGVSLCHLHSGKMEWHPTPOHTHPLLIWG 178
Db 124 FRLCGKDKLPSPVKSLEGAIVSLVSTAGTWILTHSFGLDMTWKPKAQDQPIFWGG 183
Qy 179 ATAVGQQLIQVAKHINAYTKITVTVASKKHEKLLKSYGADVDVYHDAGVIEQIKSKYPNL 238
Db 184 ATAVGQMLIQALKLNGFSKIIIVASRRKHEKLLKEYGADELFDYHDADVIEQIKKYNNI 243
Qy 239 QHVIDAVGSEDSIPEAYKVTDADSLPATLLEVVVPMTIESIPEIRKDNVKIDITLLYRAS 298
Db 244 PYLVDCVSNTEIIQQVYKCAADDATVQVLTVLTEKDIKEEDRRQNVSEGTLLYIG 303
Qy 299 QEILLGATRPASPPEYHEATVKVFKFNPHLNGDIHMMNKKVFNGLDVPALTEGIGKE 358
Db 304 NDVPFGFTLPADPEYKAAIKFINKPNKINDGQIRHPIPVRYVKNGLCDVPHILKDIK 363
Qy 359 GKNKNVYKVARL 370
Db 364 GRNSGEKLVAVL 375
RESULT 2
YL60 YEAST STANDARD; PRY; 376 AA.
AC P54007;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical 41.1 kDa protein ON CDC91-PAU4 intergenic region.
GN YLR460C OR L9122.7.
OS Saccharomycetes cerevisiae (Baker's yeast).
```

```
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OX Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
RN NCBI_TaxID=4932;
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=97313267; PubMed=9169871;
RA Johnston M., Hillier L., Riles L., Albermann K., Andre B., Ansorge W.,
RA Benes V., Brueckner M., Delius H., Dubois E., Duesterhoeft A.,
RA Entian K.-D., Floeth M., Goffeau A., Hebling U., Heumann K.,
RA Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Koetter P.,
RA Louis E.J., Messenguy F., Mewes H.-W., Miooga T., Moestl D.,
RA Mueller-Auer S., Nentwich U., Obermaier B., Piravandi E., Pohl T.M.,
RA Portetelle D., Purnelle B., Rechmann S., Rieger M., Rinke M., Rose M.,
RA Scharfe M., Scherens B., Scholler P., Schwager C., Schwarz S.,
RA Underwood A.P., Urrekarazu L.A., Vandenbol M., Verhasselt P.,
RA Vierendeels F., Voet M., Voickaert G., Voss H., Wambutt R., Wedler E.,
RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Hoheisel J.D.;
RT "The nucleotide sequence of Saccharomycetes cerevisiae chromosome XII.";
RL Nature 387:87-90(1997).
CC -!- SIMILARITY: BELONGS TO THE YCR102C/YLR460C/YNL134C FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U22383; AAB64723.1; -.
CC PIR; S59418; S59418.
CC SGD; S0004452; YLR460C.
CC InterPro; IPR002085; Adh_zn_family.
CC Pfam; PF00107; ADH_zinc_N; 1.
CC Hypothetical protein.
CC SEQUENCE 376 AA; 41127 MW; 00139949423862P1 CRC64;
CC -----
Query Match 53.2%; Score 1021.5; DB 1; Length 376;
Best Local Similarity 51.7%; Pred. No. 2.1e-66;
Matches 193; Conservative 67; Mismatches 110; Indels 3; Gaps 2;
Qy 1 MSVPTQKAVIIEGDKAVVKTDSVPELKEGTALVKVEAVAGNPTDWHKIAVKGIGESIL 60
Db 3 VAIPETMKAVIIEGDKAVVKTDSVPELKEGTALVKVEAVAGNPTDWHKIAVKGIGESIL 62
Qy 61 LGCDIAGTVVVKLGPNAS--TDLKVGDGTGFGFVHGASQTDPKNGAFAYARVYPPFYK-- 117
Db 63 LGCDAAQGVVVKLGPAVNPDKDFSIGDIYGVHIGASVRFPSNGAFAYSAISTVYKSPN 122
Qy 118 LTHSTADSEISGPKVNFESAASLPVSLTTAGVSLCHLHSGKMEWHPTPOHTHPLLIWG 177
Db 123 ELKFLGDELVPAGVRSLEGVATIPVSLTTAGVSLCHLHSGKMEWHPTPOHTHPLLIWG 182
Qy 178 GATAVQQLIQVAKHINAYTKITVTVASKKHEKLLKSYGADVDVYHDAGVIEQIKSKYPN 237
Db 183 GATAVQQLIQVAKHINAYTKITVTVASKKHEKLLKSYGADVDVYHDAGVIEQIKSKYPN 242
Qy 238 LQHVDAVGSSEDSIPEAYKVTDADSLPATLLEVVVPMTIESIPEIRKDNVKIDITLLYRAS 297
Db 243 ISYLVDCVANQDTLQQVYKCAADKQDQATIVELKNTLEENVKKNRQNVTDIIRLYSIG 302
Qy 298 QGIBILLGATRPASPPEYHEATVKVFKFNPHLNGDIHMMNKKVFNGLDVPALTEGIGKE 357
Db 303 GHEVPFGFTLPADPEYKAAIKFINKPNKINDGQIRHPIPVRYVKNGLCDVPHILKDIK 362
Qy 358 EGKKNVYKVARL 370
Db 363 YGKNSGEKLVAVL 375
RESULT 3
YCZ2 YEAST
```


RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
 RA Dunn P., Egu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 RA Giller J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Krenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
 RA Miltscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Uterback T., Van Aken S., Vayberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 thaliana.";
 RL Nature 408:816-820(2000).
 CC -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
 family. Quinone oxidoreductase subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AC005990; AAC98029.1; -;
 DR PIR; E86371; E86371.
 DR SWISS-2DPAGE; Q9ZUC1; ARATH.
 DR InterPro; IPR002085; Adh_zn family.
 DR InterPro; IPR002364; QOR_zeta_crystal.
 DR Pfam; PF00107; ADH_zinc_N; 1.
 DR PROSITE; PS01162; QOR_ZETA_CRYSTAL; FALSE_NEG.
 KW Oxidoreductase; Zinc; NAD.
 SQ SEQUENCE 309 AA; 32775 MW; EC328042771BEE6 CRC64;

Query Match 13.2%; Score 254; DB 1; Length 309;
 Best Local Similarity 32.8%; Pred. No. 3.3e-11;
 Matches 84; Conservative 35; Mismatches 97; Indels 40; Gaps 10;
 QY 15 DKAVKTDVSVPELKEGTAALVKAVAGNPTDMKHIAKYI----GPGSILGCDIAGTVV 70
 DB 13 DVLKESINIVPEIKEDQVLKIVAAALNPVDKRRQCKFKATSPLPVPGYDVGWV 72
 QY 71 KLGNASTDLKVGDTGFGFVHGASQTDPKN-GAPAEYARVVPPL--FYKSNLTHSTADEI 127
 DB 73 KVG-SAVKDLKEGDEVYANVSEKALEGPKQFGSLAEYTAVEKLLALPKNI----- 123
 QY 128 SEGPKVPKESAAISLPVSLTTAGVSLCHHLGSKMWHPTPOHTPLLIWGATVAGQOLI 187
 DB 124 -----DPAQAGLPLAETADEGLV-----RTEFSAG-----KSLIVLNGAGGVGLVI 167
 QY 188 QVAKHINAYTKIVTVASKHEKLLKSYGADDDVFDYHDAGVTEQIKSKYPNLQHVIVDAGVS 247
 DB 168 QLAHVYGASKVAATASTEKLELVRSLGADLAIDYTKEN-IEDLPDKY----DVVFDAIGM 223
 QY 248 EDS-----IPEAYKVTA 259
 DB 224 CDKAVKIKEGGKVA 239
 RESULT 6
 ID ADH2 CANAL STANDARD; PRT; 348 AA.
 AC O94038;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Alcohol dehydrogenase 2 (EC 1.1.1.1).
 GN ADH2 OR CA41C10.04.

OS Candida albicans (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
 OX NCBI_TaxID=5476;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1161;
 RA Taylor K., Harris D., Barrell B.G., Rajandream M.A.;
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +
 NADH.
 CC -1- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
 CC -1- SUBUNIT: Homotetramer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
 family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AL033501; CAA21988.1; -;
 DR PIR; T18230; T18230.
 DR InterPro; IPR002328; ADH_zinc.
 DR InterPro; IPR002085; Adh_zn family.
 DR Pfam; PF00107; ADH_zinc_N; 1.
 DR PROSITE; PS00059; ADH_ZINC; 1.
 KW Oxidoreductase; Zinc; Metal-binding; NAD.
 FT METAL 44 44 ZINC 1 (CATALYTIC) (BY SIMILARITY).
 FT METAL 67 67 ZINC 2 (BY SIMILARITY).
 FT METAL 98 98 ZINC 2 (BY SIMILARITY).
 FT METAL 101 101 ZINC 2 (BY SIMILARITY).
 FT METAL 104 104 ZINC 2 (BY SIMILARITY).
 FT METAL 112 112 ZINC 2 (BY SIMILARITY).
 FT METAL 154 154 ZINC 1 (CATALYTIC) (BY SIMILARITY).
 SQ SEQUENCE 348 AA; 36807 MW; 3D6B6753852AD2DCF CRC64;
 Query Match 9.1%; Score 175.5; DB 1; Length 348;
 Best Local Similarity 24.6%; Pred. No. 1.7e-05;
 Matches 98; Conservative 49; Mismatches 138; Indels 113; Gaps 19;
 QY 1 MSVPTTKAVITR--GDKAVVTKDVSVPKELKGTALVKVEAVGNPTD---WK---HIAY 52
 DB 1 MSVPTTKAVIFETNGKLEK-DIPVPKPKANELLINVKYSGVCHTDLHAWKGDWPLAT 59
 QY 53 KI----GPGSILGCDIAGTVVKGNGASTDLKVGDTGFGFVHGA-----SQTDPK 99
 DB 60 KLPLVGGHGG-----AGVVALGENVK-GWKVGYAGVKWLINGSCLNCEYCSQGAEPN 111
 QY 100 -----NGAFAYEARVVPPLFYKSNLTHSTADEISEGPKVKNFESAAISLPVSLTTA 148
 DB 112 CAEADLSGYTHDGSFQOYA-----TADAV-----QAARIPAGTDLA 147
 QY 149 GVS--LCCHLGSKMWHPTPOHTPLLIWGATVAGQOLIQVAKHINAYTKIVTVASKK 206
 DB 148 NVAPILCAGTVVYKALKTALEBAGVVAISGAAGGLGSLAVQYAKAM-GYRVLAIIDGGED 206
 QY 207 HEKLLKSYGADDDVFDY-HDAGVTEQIKSKYPNLQHVIVDAGVSEDSIPRAYKVTVADSLPA 264
 DB 207 KGEFVKSLGAETFDITFEKDKDVEAVKKAATNGGPHGVINVSVERAIGS----- 256
 QY 265 TLLEVVPMTIESPEIRKDNVKNIDITLLYRASGOBILLAGTRFPA-----SPEYHEATVK 320
 DB 257 -----TEYVTLGKVLVG-----LPAGAKISIPVFDVAK 287
 QY 321 FVKFINPHLNGDIHHMNIKVFNSNGLDVDPALTEGIKE 358
 DB 288 TIQIKSGYVGNRKDTAAEAVDFFTRGLIKCPKIVGLISE 325


```

RESULT 7
QOR_HUMAN
ID QOR_HUMAN STANDARD; PRT; 329 AA.
AC Q08257;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Quinone oxidoreductase (EC 1.6.5.5) (NADPH:quinone reductase) (Zeta-
DE crystallin).
DE CRYZ.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=93221534; PubMed=846529;
RA Gonzalez P., Rao P.V., Zigler J.S. Jr.;
RT "Molecular cloning and sequencing of zeta-crystallin/quinone
RT reductase cDNA from human liver.";
RL Biochem. Biophys. Res. Commun. 191:902-907(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94375054; PubMed=808825;
RA Gonzalez P., Rao P.V., Zigler J.S. Jr.;
RT "Organization of the human zeta-crystallin/quinone reductase gene
RT (CRYZ).";
RL Genomics 21:317-324 (1994).
CC -1- FUNCTION: DOES NOT HAVE ALCOHOL DEHYDROGENASE ACTIVITY. BINDS
CC NADP AND ACTS THROUGH A ONE-ELECTRON TRANSFER PROCESS.
CC ORTHOQUINONES ARE THE BEST SUBSTRATES. MAY ACT IN THE
CC DETOXIFICATION OF XENOBIOTICS.
CC -1- CATALYTIC ACTIVITY: NADPH + quinone = NADP(+) + semiquinone.
CC -1- SUBUNIT: Homotetramer.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: ONLY VERY LOW AMOUNTS IN THE LENS.
CC -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
CC family. Quinone oxidoreductase subfamily.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; L13278; AAA36536.1; -.
DR EMBL; L13526; AAK40311.1; -.
DR EMBL; L13521; AAK40311.1; JOINED.
DR EMBL; L13522; AAK40311.1; JOINED.
DR EMBL; L13523; AAK40311.1; JOINED.
DR EMBL; L13524; AAK40311.1; JOINED.
DR EMBL; L13525; AAK40311.1; JOINED.
DR PIR; P04448; P04448.
DR HSP; P28304; IQOR.
DR Genew; HGNC:2419; CRYZ.
DR MIN; 123691; -.
DR GO; GO:0007601; P:vision; TAS.
DR InterPro; IPR02085; Adh zn family.
DR Pfam; PF00107; ADH_zinc_N; 1.
DR PROSITE; PS01162; QOR_ZETA_CRYSTAL; 1.
DR QOR oxidoreductase; NADP; Zinc.
SQ SEQUENCE 329 AA; 35206 MW; 68C1828911486D4E CRC64;

Query Match 8.4%; Score 161; DB 1; Length 329;
Best Local Similarity 24.8%; Pred. No. 0.00018;
Matches 73; Conservative 50; Mismatches 95; Indels 76; Gaps 14;

```

```

QY 19 VKTDSVPELKEGTALVVKVEAVAGNPTDVKHIAKYKIGPEGSIL-----GCDIAGTVVKLGP 74
DB 24 LRSDIAPIPKDHQVLIKVHACGVNPE-TYIRSGTYSRKPLPYTPGSDVAGVIEAVGD 82
QY 75 NASTDLKVGDTGFGFVHGASQTDPKGAFAPYARVVPPLFYKSNLTHSTADEISEGPVK- 133
DB 83 NASA-FKKGRVF-----TSSTISGGVAYALAAADHTVYKL-----PEKL 121
QY 134 NFESAASLPVSLTAGVSLCHLHSGKMEWHPSTPQHTPLLIWGATAVGQQLIQVAKHI 193
DB 122 DFQGAAGIPYETAYRALHSAACVKA-----GESVLVHGASGVGLAACQIAR-- 170
QY 194 NAY-TKIV-TVASKHKEKLKSYGADVDYDHAGVIEQIKSYKPNLQHVIVDVGSDSI 251
DB 171 -AYGLKILGTAGTEGQKIVLQNGAHEVFNHREVNVDIK-KY-----VGEKG-- 217
QY 252 PEAYKTADSLPATLLEVVPMPTIESIPETRKDNVKIDITLLYRASGOEILLGA 305
DB 218 -----IDIIEMLANVNLKSLSLSH-GGRVIVWGS 248

RESULT 8
QOR_CAVPO
ID QOR_CAVPO STANDARD; PRT; 329 AA.
AC P11415;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Quinone oxidoreductase (EC 1.6.5.5) (NADPH:quinone reductase) (Zeta-
DE crystallin).
DE CRYZ.
GN Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=89378748; PubMed=2777081;
RA Rodokanaki A., Holmes R.K., Borras T.;
RT "Zeta-crystallin, a novel protein from the guinea pig lens is related
RT to alcohol dehydrogenases.";
RL Gene 78:215-224 (1989).
RN [2]
RP CHARACTERIZATION.
RC TISSUE=Liver;
RX MEDLINE=93041895; PubMed=1420281;
RA Rao P.V., Zigler J.S. Jr.;
RT "Purification and characterization of zeta-crystallin/quinone
RT reductase from guinea pig liver.";
RL Biochim. Biophys. Acta 1117:315-320(1992).
RN [3]
RP CHARACTERIZATION.
RC TISSUE=Liver;
RX MEDLINE=92112732; PubMed=1370456;
RA Rao P.V., Krishna C.M., Zigler J.S. Jr.;
RT "Identification and characterization of the enzymatic activity of
RT zeta-crystallin from guinea pig lens. A novel NADPH:quinone
RT oxidoreductase.";
RL J. Biol. Chem. 267:96-102(1992).
CC -1- FUNCTION: DOES NOT HAVE ALCOHOL DEHYDROGENASE ACTIVITY. BINDS
CC NADP AND ACTS THROUGH A ONE-ELECTRON TRANSFER PROCESS.
CC ORTHOQUINONES ARE THE BEST SUBSTRATES. MAY ACT IN THE
CC DETOXIFICATION OF XENOBIOTICS.
CC -1- CATALYTIC ACTIVITY: NADPH + quinone = NADP(+) + semiquinone.
CC -1- SUBUNIT: Homotetramer.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: IN GUINEA PIGS IT CONSTITUTES ABOUT 10% OF THE
CC WATER SOLUBLE PROTEINS OF THE LENS.
CC -1- DISEASE: A GENOMIC MUTATION CAUSING THE DELETION OF 34 AMINO ACIDS
CC WAS CLEARLY ASSOCIATED WITH A HEREDITARY NUCLEAR CATARACT IN A
CC LINE OF STRAIN 13 GUINEA PIGS.
CC -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
CC family. Quinone oxidoreductase subfamily.

```

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 RA EMBL; M26936; AAA37035.1; -;
 RA PIR; J50230; CYGEPZ.
 RA HSSP; P28304; 1QOR.
 RA InterPro; IPR002085; Adh zn family.
 RA InterPro; IPR002364; QOR_zeta_crystal.
 RA Pfam; PF00107; ADH_zinc_N; 1.
 RA PROSITE; PS01162; QOR_ZETA_CRYSTAL; 1.
 KW Oxidoreductase; NADP; Zinc; Eye lens protein.
 SQ SEQUENCE 329 AA; 35202 MW; 1463632C572C234F CRC64;

 Query Match 8.3%; Score 159; DB 1; Length 329;
 Best Local Similarity 24.2%; Pred. No. 0.00025;
 Matches 71; Conservative 50; Mismatches 96; Indels 76; Gaps 14;

 QY 19 VKTDSVPELKEGTALVKVEAVAGNPTDWRKHAIYKIGPEGSL-----GCDIAGTVVVLGP 74
 DB 24 VQSDVAVPIPKDHQVLIKVKHACGINPVE-TVIRSGTYTRIPLLPYTPGTDVAGVVEISGN 82
 QY 75 NASTDLKVGDTGFGFVHGASQTDPKNGAFAPARYVPPFPYKSNLTHSTADEISEGPKV- 133
 DB 83 DVSA-FKKGDRVF-----TTSTISGAYAEYALASDHTVRL-----PEKL 121
 QY 134 NFESAASLPVSLTTAGVSLCHLHGSKMEWHPSTPQHTPLLLMGATVAGQQLIOVAKHI 193
 DB 122 DFRQALGIPYFTA-----CRAL-----FHSARAKAGESVLVHGASGVGLAACQIAR-- 170
 QY 194 NAY-TKIV-TVASKKHEKLSYGDVDFVDHAGVIEQIKSKYPNLQHVIVDAVGSSEDSI 251
 DB 171 -AYGLKVLGTAGTEEGQVVLONGAHEVFNHRDAHYDEIKK-----SIGEGK-- 217
 QY 252 PEAYKVTADSLPVLLEVPVFTIESPEIRKONKVIDITLLYRASQGEILLG 304
 DB 218 -----VDVIEMLANVNLNDKLL-SCGGRVIVG 247

 RESULT 9
 QOR_MOUSE STANDARD; PRT; 331 AA.
 AC P47199; Q62508; Q99L63;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Quinone oxidoreductase (EC 1.6.5.5) (NADPH:quinone reductase) (Zeta-crystallin).
 DE Crystallin.
 GN CRYZ.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OK NCBI_TaxID=10090;
 RN [1]
 RN TISSUE=Liver;
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=94224126; PubMed=8170370;
 RA Gonzalez P., Hernandez-Calzadilla C., Rao P.V., Rodriguez I.R., Ziegler J.S. Jr., Borras T.;
 RT "Comparative analysis of the zeta-crystallin/quinone reductase gene in guinea pig and mouse."
 RL Mol. Biol. Evol. 11:305-315(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Breast tumor;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Scapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Uedlin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahney J., Heiton E., Kettman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinaki M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP SEQUENCE OF 130-331 FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Brain cortex;
 RX MEDLINE=96216731; PubMed=8645260;
 RA Kajiwara K., Nagasawa H., Shimizu-Nishikawa K., Ookura T., Kimura M., Sugaya E.;
 RT "Molecular characterization of seizure-related genes isolated by differential screening."
 RL Biochem. Biophys. Res. Commun. 219:795-799(1996).
 CC -/- FUNCTION: DOES NOT HAVE ALCOHOL DEHYDROGENASE ACTIVITY. BINDS NADP AND QUINONE THROUGH A ONE-ELECTRON TRANSFER PROCESS. ORTHOQUINONES ARE THE BEST SUBSTRATES. MAY ACT IN THE DETOXIFICATION OF XENOBIOTICS (BY SIMILARITY).
 CC -/- CATALYTIC ACTIVITY: NADPH + quinone = NADP(+) + semiquinone.
 CC -/- SUBUNIT: Homotetramer.
 CC -/- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -/- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase family. Quinone oxidoreductase subfamily.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 DR EMBL; S70056; AAB30620.2;
 DR EMBL; BC003800; AAH03800.1; -;
 DR EMBL; D78646; BAA11463.1; -;
 DR PIR; A54932; A54932.
 DR HSSP; P28304; 1QOR.
 DR MGD; MGI:88527; Crz.
 DR InterPro; IPR002085; Adh_zn_family.
 DR InterPro; IPR002364; QOR_zeta_crystal.
 DR Pfam; PF00107; ADH_zinc_N; 1.
 DR PROSITE; PS01162; QOR_ZETA_CRYSTAL; 1.
 KW Oxidoreductase; NADP; Zinc.
 FT CONFLICT 58 58 A -> T (IN REF. 2).
 FT CONFLICT 131 133 IPY -> TMD (IN REF. 3).
 SQ SEQUENCE 331 AA; 35268 MW; 35816C043EFE16A2 CRC64;

 Query Match 8.3%; Score 158.5; DB 1; Length 331;
 Best Local Similarity 22.4%; Pred. No. 0.00027;
 Matches 72; Conservative 53; Mismatches 119; Indels 77; Gaps 13;

 QY 19 VKTDSVPELKEGTALVKVEAVAGNPTDWRKHAIYKIGPEGSL-GCDIAGTVVVLGP 75
 DB 24 VQSDVAVPIPKDHQVLIKVKHACGINPVE-TVIRSGTYTRIPLLPYTPGTDVAGVVEISGVGDK 83
 QY 76 ASTDLKVGDTGFGFVHGASQTDPKNGAFAPARYA-----RVPPFLYKSNLTHSTADEISEG 130
 DB 84 VSA-FKKGDRVFCY-----STVSGGYAEFALADDTIYP-----LPE- 119
 QY 131 PVKNFESAASLPVSLTTAGVSLCHLHGSKMEWHPSTPQHTPLLLMGATVAGQQLIOVA 190

```
Db 120 -TLNFRGNAALGIPYFTA-----CRAL-----FHSARARAGESVLVHGASGVGLATQCIA 169
Qy 191 KHINAYTKIVTASKEKELKSKGADDPYDHDAGVIEQIKSKYPNLQHVDAVGSSEDS 250
Db 170 R-ANGLKVLGTAGSEBEGKVLQNGAHEVFNKKAINDIKM-----SVGDKK 218
Qy 251 IPEAYKVTADSLPATLLEWVPMTIESIPBEIRKONKVIDITLLYRAGQBIILLGATRPFA 310
Db 219 -----GVDVITIELMANENLNDLKL--SHGGRVVVVGCRGPI 254
Qy 311 SPEYHEATVKPVKEINPHLNN 331
Db 255 EINPRDTWAKETSIIGVSLSS 275

RESULT 10
QD_LEIAM
ID_QD_LEIAM STANDARD; PRT; 340 AA.
AC P42865;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Possible quinone oxidoreductase (EC 1.6.5.5) (NADPH:quinone
DE reductase) (P36).
OS Leishmania amazonensis.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OC NCBI_TaxID=5659;
OX [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=95107343; PubMed=7808470;
RA Liu X., Chang K.P.;
RT "Identification by extrachromosomal amplification and overexpression
RT of a zeta-crystallin/NADPH-oxidoreductase homologue constitutively
RT expressed in Leishmania spp.";
RL Mol. Biochem. Parasitol. 66:201-210(1994).
CC -1- CATALYTIC ACTIVITY: NADPH + quinone = NADP(+) + semiquinone.
CC -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
CC family. Quinone oxidoreductase subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L11705; AAA73554.1; -
DR InterPro; IPR002085; Adh zn family.
DR InterPro; IPR002364; QOR_zeta_crystal.
DR Pfam; PF00107; Adh zinc N; 1.
DR PROSITE; PS01162; QOR_ZETA_CRYSTAL; 1.
KW Oxidoreductase; NADP; Zinc.
SQ SEQUENCE 340 AA; 36272 MW; 4D9255AB56856625 CRC64;

Query Match 8.1%; Score 155; DB 1; Length 340;
Best Local Similarity 23.5%; Pred. No. 0.00051;
Matches 92; Conservative 58; Mismatches 151; Indels 90; Gaps 19;

Qy 1 MSVPTTKAVIIEGDKAVKVTDSV-----PELKEGTALVKVEAVAGNPTDMKHAIYKI 54
Db 1 MSSPSNFKKLVSLSKDFRSSTSVVEAHLPEEVEPGNVRVSVKYAGVNASDLNF----- 55
Qy 55 GPESGIL-----GCDINGTVVKVGLPNASTDLKVGDTGFGFVGASQTDPKNGAPAE 105
Db 56 -TNGSYFNKVNQPPFDGCFEAGTGVQIGAGVA-NVKVGD-----HVVLM---QYGCFAE 104
Qy 106 YARVVPPLFYKSLNLTHTSADSEISG--PVKNFESAAS-LPVSLLTAGVSLCHLGSXMEW 162
Db 105 F-----LDAPERCIPVPELKPESVLPVPSAULTAAVAL-GEVGRVKKG 146
Qy 163 HPSTPQHTHPLLINGGATAVGQQLIQVAKHINAYTKIVTASKEKELKSKGADDPVFDY 222
```

```
Db 147 DVA-----LVTAAAGGTGQIAVQLLKHVYCTVIGTSSSEKAFLKSGICDHVINY 198
Qy 223 HDAGVIEQIKSKYP-NLQHVDAVGSSEDSIPRAYKVTADSLPATL-----EVVPM 272
Db 199 KTESLDRHLHELCPKGVVYECVGGH-TFNDARVHVAVHARVVIIIGSTSSYSKSGEVVPP 257
Qy 273 TIESIPBEIRKONKVIDITLLYR-ASGQEIILLGATRPFAPEYHEATVKPVKEINPHLNN 331
Db 258 SDPS-----GTSVTMLLLVKASLNGFFL-----POPHDVI PKYMANLLQYLKA 301
Qy 332 GDTH-HMNVKVFNSGLDDVPALTEGKEGKN 361
Db 302 GQVKLFVDKKVF-HGLSSVADAVDHLVSGAN 331

RESULT 11
YM27_YEAST
ID_YM27_YEAST STANDARD; PRT; 365 AA.
AC P28635;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical 41.6 kDa protein in IMPI-HLJI intergenic region (RF1095).
DE YMR152W OR YMR375.22 OR YMR520.01.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=91360060; PubMed=1886606;
RA Bahrens M., Michaelis G., Prati E.;
RT "Mitochondrial inner membrane protease 1 of Saccharomyces cerevisiae
RT shows sequence similarity to the Escherichia coli leader peptidase.";
RL Mol. Gen. Genet. 228:167-176(1991).
RN [2]
RN SEQUENCE FROM N.A.
RX STRAIN=S288c / AB972;
RX PubMed=9169872;
RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA Jagels K., Lyne G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT XIII.";
RL Nature 387:90-93(1997).
CC -1- SIMILARITY: SOME, TO YEAST AST1/AST2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S55518; AAB19702.1; -
DR EMBL; Z47071; CAA87367.1; -
DR EMBL; Z49705; CAA89788.1; -
DR PIR; S50409; S50409.
DR SGD; S0004760; YIM1.
DR GO; GO:0005811; C:lipid particle; IDA.
DR GO; GO:0005743; C:mitochondrial inner membrane; IDA.
DR GO; GO:0008233; F:peptidase activity; IMP.
DR GO; GO:0006627; P:mitochondrial processing; IMP.
DR InterPro; IPR002085; Adh zn family.
DR Pfam; PF00107; Adh zinc N; 1.
KW Hypothetical protein; Transmembrane
TRANSMEM 140 158 POTENTIAL.
FT CONFLICT 121 121 V -> I (IN REF. 1).
FT CONFLICT 234 234 S -> I (IN REF. 1).
FT CONFLICT 280 280 D -> N (IN REF. 1).
```

```
FT CONFLICT 296 296 L -> S (IN REF. 1).
SQ SEQUENCE 365 AA; 41637 MW; 88F6453D9E918A16 CRC64;

Query Match 8.0%; Score 154; DB 1; Length 365;
Best Local Similarity 23.5%; Pred. No. 0.00066;
Matches 63; Conservative 49; Mismatches 116; Indels 40; Gaps 12;

QY 6 TOKAVIIBGDKAVKVTDSVPPEL-----EGTALVKVAVAGNPTDW-KH---IAYKIGPE 57
DB 7 TNKSVTYVNTPTVITISSELDLSCYQDDVEVHAAALNPIDFTHQLCNSYIFGKY 66
QY 58 GSILGCDIAGTVKVLGPNASTDLKVGDTGFGF-----VHGASQTDPKNGAFAYARVYP--- 111
DB 67 PKTYSRDSYGVIIKAGKDQVNRKVGDKVNGWYSHVIG-----ERGLTHYLILNPAKD 120
QY 112 -PLPKSNLTHSTADEISEGPKVPFESAASLPVSLTTAGVSLCHLGSKMEHWPSTPQHT 170
DB 121 VPI---THMVRVPKDE--NDPYDDFVYAAAWPLTFTGAFSTL---YDFKKDWTSDS---- 168
QY 171 HPLIIGGATAGVQGLIQVAKHINAYTKIVTVASKKHEKLKSYGADDPVDYHDAGVIEQ 230
DB 169 -KVLVIGASTSVSAFVHIAKNYFNIGTVGICNSIERNKGLGYDYLVPYDEGSIVEN 227
QY 231 IKS-KYPNLOH-----VIDAVGSEDSIP 252
DB 228 VKLKQSVLENDKFDIMFDSVGNHDFRP 255

RESULT 12
P1 ARATH
ID P1 ARATH STANDARD; PRT; 345 AA.
AC Q35172;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable NADP-dependent oxidoreductase P1 (EC 1.1.1.1).
GN P1 OR AT5G16970 OR P2K13 120.
OS Arabidopsis thaliana (Mouae-ear creas).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=96064691; PubMed=7592828;
RA Babylchuk E., Kushnir S., Belles-Boix E., van Montagu M., Inze D.;
RT "Arabidopsis thaliana NADPH oxidoreductase homologs confer tolerance
of yeasts toward the thiol-oxidizing drug diamide.";
RL J. Biol. Chem. 270:26224-26231(1995).
RN [2];
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016721; PubMed=11130714;
RA Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,
RA Miyajima N., Sasaoka S., Kimura T., Hosouchi T., Kawashima K.,
RA Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,
RA Watanabe N., Naruo K., Okumura S., Shinpo S., Takeuchi C., Wada T.,
RA Katsunuma N., Yanada M., Yasuda M., Sato S., de la Bastide M.,
RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,
RA Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,
RA Stoneking T., Pepin K., Spieth J., Sekhon M., Armetrong J., Becker M.,
RA Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,
RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,
RA Leonard S., Meyer R., Mulvaney E., Ozerky P., Riley A., Strommatt C.,
RA Wagner-McPherson C., Wollam A., Yoakum M., Bell M., Dedhia N.,
RA Parneill L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,
RA Kirchoff K., Toth K., King L., Bahret A., Miller B., Marra M.,
RA Martensen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
RA Volckaert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T.,
RA Entian K.-D., Terryn N., Hartley N., Bent E., Johnson S.,
RA Langham S.-A., McCullagh B., Robben J., Grymonprez B., Zimmermann W.,
RA Ransperger U., Wedler H., Balke K., Wedler E., Peters S.,
```

```

Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
NCBI_TaxID=4911;
SEQUENCE FROM N.A.
STRAIN=ATCC 12424;
MEDLINE=93250057; PubMed=8485163;
Ladriere J.M., Delcour J., Vandenhaute J.;
"Sequence of a gene coding for a cytoplasmic alcohol dehydrogenase
from Kluyveromyces marxianus ATCC 12424.";
Biochim. Biophys. Acta 1173:99-101(1993).
-1- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +
NADH
-1- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
-1- SUBUNIT: Homotetramer.
-1- SUBCELLULAR LOCATION: Cytoplasmic.
-1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X60224; CAA42785.1; -.
PIR; S32521; S32521.
InterPro; IPR002328; ADH_zinc.
InterPro; IPR002085; Adh_zn_family.
Pfam; PF00107; ADH_zinc_N; 1.
PROSITE; PS00059; ADH_ZINC; 1.
Oxidoreductase; Zinc; Metal-binding; NAD; Multigene family.
METAL 44 44 ZINC 1 (CATALYTIC) (BY SIMILARITY).
METAL 67 67 ZINC 1 (CATALYTIC) (BY SIMILARITY).
METAL 98 98 ZINC 2 (BY SIMILARITY).
METAL 101 101 ZINC 2 (BY SIMILARITY).
METAL 104 104 ZINC 2 (BY SIMILARITY).
METAL 112 112 ZINC 2 (BY SIMILARITY).
METAL 154 154 ZINC 1 (CATALYTIC) (BY SIMILARITY).
SEQUENCE 348 AA; 37158 MW; A75D2EBE82E355BD CRC64;
Query Match 8.0%; Score 153.5; DB 1; Length 348;
Best Local Similarity 22.6%; Pred. No. 0.0067;
Matches 90; Conservative 53; Mismatches 142; Indels 113; Gaps 20;
Qy 1 MSVPTTKAVII--EGDKAVVKTDSVPVPELKEGTALVKVE-----AVAGN-PTDW 47
Db 1 MAIPETQKGVIFYEHGSGSLQYK-DIPVKKPNELLINNVKSYGVCHTDLHAWQGDWPLDT 59
Qy 48 KHIAVKIGPEGSILGCDIAGTVVKLGPNASTDLKVG-D-TGFGFVHGASQT-----96
Db 60 K-LPLVGGHEG-----AGIVVANGENV-TGWEIGDVAGIKWLNGSCMCECELSNEP 110
Qy 97 -DPK-----NGAFAYARYVPPLFYKSNLTHSTADEISEGPV-KNFESAASLPVSLT 146
Db 111 NCPRADLSGYTHDGSFQYA-----TADAVQAARIPKNVDAEVAPI--L 153
Qy 147 TAGVSLCHLGSKM-----EWPSTPQTHPLLWGATAVGQQQLQVAKHINATKIVTV 202
Db 154 CAGVTYVYKALKSAHIKAGDW-----VAISGACGGLSLAIQYAKAN-GYRVLGID 202
Qy 203 ASKHKELKLSYGADDDVFDYHDAGVIEQIKSKPNLQHVDAVSGSDSIPKAYKVTADSL 262
Db 203 AGDEKAKLFKELG-----EYFDFTKTKQNVAEIVATNGVA 240
Qy 263 PATLLEVVPMTIESIPBEIRKDNVKIDITLTY-RASGOEILLGATR-PPASPEYHEATVK 320
Db 241 HAVI-----NVSVSEAAISTSVLYTRSGTVLVGLPRDAQCKSDVFNQVVK 287
Qy 321 FVKFINPLHNGDIIHNMKIVFSNGLDDVDPALITGRIKE 358
Db 288 SISIVGSYVGNRADTREALDFFSRGLVKAPIKILGLSE 325
-----
RESULT 14
ADH2_PICST STANDARD; PRT; 348 AA.
AC O13309; O00090;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Alcohol dehydrogenase II (EC 1.1.1.1) (ADH 1).
GN ADH2 OR ADH1.
OS Pichia stipitis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.
OX NCBI_TaxID=4924;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 58785 / CBS 6054;
RX MEDLINE=98207839; PubMed=9546172;
RA Cho J.Y., Jeffries T.W.;
RT "Pichia stipitis genes for alcohol dehydrogenase with fermentative
and respiratory functions.";
RL Appl. Environ. Microbiol. 64:1350-1358(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS 5774;
RX MEDLINE=99019018; PubMed=9802210;
RA Passoth V., Schaefer B., Liebel B., Weierstall T., Kliner U.;
RT "Molecular cloning of alcohol dehydrogenase genes of the yeast Pichia
stipitis and identification of the fermentative ADH.";
RL Yeast 14:1311-1325(1998).
CC -1- FUNCTION: CONVERTS ETHANOL TO ACETALDEHYDE.
CC -1- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +
NADH.
CC -1- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
CC -1- SUBUNIT: Homotetramer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AF008244; AAC49990.1; -.
EMBL; Y13238; CAA73690.1; -.
InterPro; IPR002328; ADH_zinc.
InterPro; IPR002085; Adh_zn_family.
Pfam; PF00107; ADH_zinc_N; 1.
PROSITE; PS00059; ADH_ZINC; 1.
Oxidoreductase; Zinc; Metal-binding; NAD; Multigene family.
METAL 44 44 ZINC 1 (CATALYTIC) (BY SIMILARITY).
METAL 67 67 ZINC 1 (CATALYTIC) (BY SIMILARITY).
METAL 98 98 ZINC 2 (BY SIMILARITY).
METAL 101 101 ZINC 2 (BY SIMILARITY).
METAL 104 104 ZINC 2 (BY SIMILARITY).
METAL 112 112 ZINC 2 (BY SIMILARITY).
METAL 154 154 ZINC 1 (CATALYTIC) (BY SIMILARITY).
CONFLICT 41 41 MISSING (IN REF. 2).
SEQUENCE 348 AA; 35565 MW; F6C813C98F56A148 CRC64;
Query Match 7.9%; Score 152.5; DB 1; Length 348;
Best Local Similarity 24.3%; Pred. No. 0.00079;
Matches 97; Conservative 51; Mismatches 136; Indels 115; Gaps 21;
Qy 1 MSVPTTKAVIIE--GDKAVVKTDSVPVPELKEGTALVKVEAVAGNPTD---WK-----48
Db 1 MSIPPTKAVIFETNGGPLYK-DIPVKKPNELLINNVKSYGVCHTDLHAWQGDWPLDT 59

```

```
QY 49 HIAYKIGPEGSILGCDIAGTVVVKLGNASTDLKVGDD-TGFGFVHG-----ASQTD 97
DQ 60 KLPLVGGHEG-----AGVVVALGENV-TGWEIGDYGAKWINGSCLQCEYCVTAHESN 111
QY 98 -----PKNGAFARVYVPLFYKSNLTHSTADEISEGPV-KNFESAASLPVSLTT 147
DQ 112 CPDADLSGYTHDGSFQQA-----TADAIAQAARIPKGTDLALAPI--LC 154
QY 148 AGVSLCHHLGSKM-----EMHPSTQHTPLLIWGATVGOOLIOVAKHINAYTKIVTA 203
DQ 155 AGITVYKALKTAQLOAGQW-----VAVSAGAGLGLSLAIQYAKAMGYRVVGDGG 204
QY 204 SKHEKLLKSYGADDVFDYHDAGVIEQIKSKYPNLQHVDAVGSSEDSIPEAYKVTADSLP 263
DQ 205 ADKGE-FAKSLGA-EVP-----VDFLSKD-----VADVILK 234
QY 264 ATLEVVPMVTIESIPEIRKDNKVIDITLLYRASGOBILGATPPA-----SPHYHATV 319
DQ 235 ATNGAGHGVINVSERAMQOSVD-----YVRPTGTVVLVG---LPAGAKVSASVSSVV 286
QY 320 KFKVFINPLNNGDIHNMNIKVFNSGLDDVPALTEGIKE 358
DQ 287 RTIQIKSGVGNRADSAEADFFTRGLIKCPKIVVGLSE 325

RESULT 15
ADHL_PICST
ID_ADHL_PICST STANDARD; PRT; 348 AA.
AC 000097;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Alcohol dehydrogenase I (EC 1.1.1.1) (ADH 2).
GN ADH1 OR ADH2.
OS Pichia stipitis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.
OX NCBI_TaxID=4924;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 58785 / CBS 6054;
RX MEDLINE=98207839; PubMed=9546172;
RA Cho J.Y., Jeffries T.W.;
RT "Pichia stipitis genes for alcohol dehydrogenase with fermentative
and respiratory functions.";
RL Appl. Environ. Microbiol. 64:1350-1358(1998).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=CBS 5774;
RX MEDLINE=95019018; PubMed=9802210;
RA Passoth V., Schaefer B., Liebel B., Weierstall T., Klinger U.;
RT "Molecular cloning of alcohol dehydrogenase genes of the yeast Pichia
stipitis and identification of the fermentative ADH.";
RL Yeast 14:1311-1325(1998).
CC -1- FUNCTION: CONVERTS ETHANOL TO ACETALDEHYDE AND PLAYS A MAJOR ROLE
IN XYLOSE FERMENTATION.
CC -1- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +
NADH.
CC -1- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
CC -1- SUBUNIT: Homotetramer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/
or send an email to license@isb-sib.ch).
-----
EMBL; AF008245; AAC49991.1; -.
```

```
DR EMBL; Y13397; CAA73827.1; -.
DR InterPro; IPR002328; ADH_zinc.
DR InterPro; IPR002085; Adh_zn_family.
DR Pfam; PF00107; Adh_zinc_N; 1.
KW PROSITE; PS00059; ADH_ZINC; 1.
KW Oxidoreductase; Zinc; Metal-binding; NAD; Multigene family.
FT METAL 44 44 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 67 67 ZINC 2 (BY SIMILARITY).
FT METAL 98 98 ZINC 2 (BY SIMILARITY).
FT METAL 101 101 ZINC 2 (BY SIMILARITY).
FT METAL 104 104 ZINC 2 (BY SIMILARITY).
FT METAL 112 112 ZINC 2 (BY SIMILARITY).
FT METAL 154 154 ZINC 1 (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 348 AA; 36520 MM; 49C06B545D5350F4 CRC64;
Query Match 7.8%; Score 149.5; DB 1; Length 348;
Best Local Similarity 23.6%; Pred. No. 0.0013;
Matches 92; Conservative 51; Mismatches 150; Indels 97; Gaps 18;
QY 1 MSVPTTQKAVIIEGDKA-VVKTDVSVPELKEGTALVKVEAVAGNPTD---WK-----H 49
DQ 1 MSVPTTQKAVFESNGGPLYKDIPTPKPNEILINVKYSGVCHTDLHAWKGDWLDTK 60
QY 50 IAYKIGPEGSILGCDIAGTVVVKLGNASTDLKVGDD-TGFGFVHG-----SQTDPPNG 101
DQ 61 LPLVGGHEG-----AGVVVGIGSNV-TGWEIGDYAGIKWINGSCLNCFQHSDEPNC 112
QY 102 APAYEARVPPPLFYKSNLTH-----STADEISEGPVKNFESAASLP--VSLTTAGVSL 152
DQ 113 AKADL-----SGYTHDGSFQOYATADAV-----QAARLPKGTDLAQAPIL 153
QY 153 CHHLGSKMEHPSTPOHTPLLIWGATVGOOLIOVAKHINAYTKIVTVASKKHEKLLK 212
DQ 154 CAGITVYKALKTAQIQGNWVCISGAGGLGSLAIQYAKAM-GFRVIAIDGGEKEGFVK 212
QY 213 SYGADVDVFDYHDAGVIEQIKSKYPNLQHVDAVGSSEDSIPEAYKVTADSLPALLEVPM 272
DQ 213 SLGAEAVVDF-----TVSKDIVKDIQTAT-DGSPHAAINV--- 246
QY 273 TIESIPEIRKONKVIDITLLYRASGOBILGATPPASPE-----YHEATVKKVKEFINPH 328
DQ 247 ---SVSEKAIAOSQ-----YVRSTGTVVLVG---LPAGAKVVAVPFDVAVKVSISIRGSY 295
QY 329 LNNGDIHNMNIKVFNSGLDDVPALTEGIKE 358
DQ 296 VGNRADSAEADFFTRGLIKCPKIVVGLSE 325

RESULT 16
QOR_BOVIN
ID_QOR_BOVIN STANDARD; PRT; 330 AA.
AC Q97784;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zeta-crystallin.
GN CRYZ.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Lens;
RX MEDLINE=97299768; PubMed=9154917;
RA Rao P.V., Gonzalez P., Persson B., Joernvall H., Garland D.,
Zigler J.S. Jr.;
RT "Guinea pig and bovine zeta-crystallins have distinct functional
characteristics highlighting replacements in otherwise similar
structures.";
RL Biochemistry 36:5353-5362(1997).
CC -1- FUNCTION: DOES NOT HAVE QUINONE REDUCTASE ACTIVITY. BINDS STRONGLY
```



```
CC TO SINGLE-STRANDED DNA.
CC -1- SUBUNIT: Homotetramer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
CC family. Quinone oxidoreductase subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U70048; RAD10290.1; -.
DR HSSP; P28304; IQOR.
DR InterPro; IPR002085; Adh_zn_family.
DR Pfam; PF00107; ADH_zinc_N; 1.
DR PROSITE; PS01162; QOR_ZETA_CRYSTAL; 1.
KW Zinc; DNA-binding.
SQ SEQUENCE 330 AA; 35403 MW; BD24ACD857BA8673 CRC64;

Query Match 7.7%; Score 148; DB 1; Length 330;
Best Local Similarity 24.0%; Pred. No. 0.0016;
Matches 71; Conservative 54; Mismatches 91; Indels 80; Gaps 15;

Qy 19 VKTDSVPELKEGTALVKVEAVAGNPTDM--KHAYKIGPEGSTL---GCDIAGTVVKL 72
Db 24 LQSDVAVPIPEKHQVLKQACGVNPDVTYIRSTHNIKP---LLPTPGDFVAGIIEAV 80
Qy 73 GPNASTDLKVGDTGFGFVHGASQTDPKNGAFAYRVPYPLFYKSNLTHSTADEISEGPV 132
Db 81 GESVSA-FKKGRVF-----TTRTISGGVAYEALAAADHTVYL-----PE 119
Qy 133 K-NFESAASLPVSLTTAGVSLCHLHSGKMEWHPTPOHTPLLIWGATVAGQOLIIVAK 191
Db 120 KLDFKQGAAGIPYFTAYRALLYSY-----PVKPGES--VLVHGASGGVGIAACQIAR 170
Qy 192 HINAY-TKIV-TVASKKHKLKSYGADDVFDYHDAGVIEQIKSYKNLQHVDAVGSSED 249
Db 171 ---AYGLKVLGTSTEGQKIVLENGAKVFNHKEANYIDIKKK-----SVGEKG 217
Qy 250 SIPEAYKVTDASLPATLLEVPVMTIESIPEIRKDNVKIDITLLYRASGOEILGA 305
Db 218 -----VDVIEMLANVNLNDLNLSSH--GGRVIVVGS 248

RESULT 17
QOR_LAMGU STANDARD; PRT; 330 AA.
AC Q28452;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE Quinone oxidoreductase (EC 1.6.5.5) (NADPH:quinone reductase) (Zeta-
DE crystallin).
GN CRYZ.
OS Lama guanicoe (Guanaco).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Lama.
OX NCBI_TaxID=9840;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lens;
RX MEDLINE=96073447; PubMed=7476124;
RA Gonzalez P., Rao P.V., Nunez S.B., Zigler J.S. Jr.;
RT "Evidence for independent recruitment of zeta-crystallin/quinone
RT reductase (CRYZ) as a crystallin in camels and hystriomorph
RT rodents."
RL Mol. Biol. Evol. 12:773-781(1995).
CC -1- FUNCTION: DOES NOT HAVE ALCOHOL DEHYDROGENASE ACTIVITY. BINDS
CC NADP AND ACTS THROUGH A ONE-ELECTRON TRANSFER PROCESS.
```

```
CC ORTHOQUINONES ARE THE BEST SUBSTRATES. MAY ACT IN THE
CC DETOXIFICATION OF XENOBIOTICS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: NADPH + quinone = NADP(+) + semiquinone.
CC -1- SUBUNIT: Homotetramer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
CC family. Quinone oxidoreductase subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L34159; AAA99986.1; -.
DR HSSP; P28304; IQOR.
DR InterPro; IPR002085; Adh_zn_family.
DR Pfam; PF00107; ADH_zinc_N; 1.
DR PROSITE; PS01162; QOR_ZETA_CRYSTAL; 1.
KW Oxidoreductase; NADP; Zinc.
SQ SEQUENCE 330 AA; 35187 MW; E784E414D2BA23D6 CRC64;

Query Match 7.7%; Score 148; DB 1; Length 330;
Best Local Similarity 23.1%; Pred. No. 0.0016;
Matches 67; Conservative 51; Mismatches 104; Indels 68; Gaps 11;

Qy 19 VKTDSVPELKEGTALVKVEAVAGNPTDM--KHAYKIGPEGSTL---GCDIAGTVVKLGN 75
Db 24 LQSDVAVPIPEKHQVLKQACGVNPDVTYIRSTGYSRKPRLPYTGLDVAIGLIEAVGR 83
Qy 76 ASTDLKVGDTGFGFVHGASQTDPKNGAFAYRVPYPLFYKSNLTHSTADEISEGPVKNF 135
Db 84 VSA-FKKGRVF-----TTRTISGGVAYEALAAADHTVYL-----LPGL 123
Qy 136 ESAASLPVSLTTAGVSLCHLHSGKMEWHPTPOHTPLLIWGATVAGQOLIIVAKHINA 195
Db 124 QKGAIGVPYFTAYRALLHSACA-----GESVLHGASGGVGLAACQIARAC-C 173
Qy 196 YTKIVTVASKKHKLKSYGADDVFDYHDAGVIEQIKSYKNLQHVDAVGSSEDSIPEAY 255
Db 174 FKVLGTAGTEGQRVVLQNGAHEVFNHREDINIDIKKK-----SVGEKG----- 217
Qy 256 KVTADSLPATLLEVPVMTIESIPEIRKDNVKIDITLLYRASGOEILGA 305
Db 218 -----IDVIEMLANVNLNDLNLSSQ--GGRVIVVGS 248

RESULT 18
P2_ARATH STANDARD; PRT; 343 AA.
AC Q39173; Q9LFK3;
DT 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable NADP-dependent oxidoreductase P2 (EC 1.-.-.-).
GN P2 OR AT5G16990 OR F2K13_140.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=96064691; PubMed=7592828;
RA Babychuk E., Kuehnir S., Belles-Boix E., van Montagu M., Inze D.;
RT "Arabidopsis thaliana NADPH oxidoreductase homologs confer tolerance
RT of yeasts toward the thiol-oxidizing drug diamide."
RL J. Biol. Chem. 270:26224-26231(1995).
RN [2]
```

SEQUENCE FROM N.A.
 RC STRAIN=cv, Columbia;
 RX MEDLINE=21016721; PubMed=11130714;
 RA Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,
 RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,
 RA Kohara M., Matsumoto M., Matsumoto A., Muraki A., Nakayama S.,
 RA Nakazaki N., Naruo K., Okumura S., Shino S., Takeuchi C., Wada T.,
 RA Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,
 RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,
 RA Stemann K., Murray J., Johnson D., Rohlfing T., Nelson J.,
 RA Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,
 RA Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,
 RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,
 RA Leonard S., Meyer R., Mullaney E., Ozersky P., Riley A., Strommatt C.,
 RA Wagner-McPherson C., Wollam A., Yoakum M., Bell M., Dedhia N.,
 RA Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,
 RA Kirchhoff K., Toth K., King L., Bahret A., Miller B., Marra M.,
 RA Martinsson R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
 RA Volckaert G., Wambutt R., Duesterhoef A., Stiekema W., Pohl T.,
 RA Entian K.-D., Terry N., Hartley N., Bent E., Johnson S.,
 RA Langham S.-A., McCullagh B., Robben J., Grymonprez B., Zimmermann W.,
 RA Ramsperger U., Wedler H., Baake K., Wedler E., Peters S.,
 RA van Staveren M., Dirkse W., Mooljman P., Klein Lankhorst R.,
 RA Weitzenecker T., Bothe G., Rose M., Hauf J., Berneiser S., Hempel S.,
 RA Feldpausch M., Lamberth S., Villarroel R., Giesen J., Ardiles W.,
 RA Bents O., Lemcke K., Kolesov G., Mayer K.F.X., Rudd S., Schoof H.,
 RA Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Franz P.F.;
 RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis
 RT thaliana.";
 RL Nature 408:823-826(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv, Columbia;
 RA Shinzaki K., Davis R.W., Ecker J.R., Theologis A.;
 RT "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
 RT SSP consortium (Salk/Stanford/PGSC).";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MAY PLAY A DISTINCT ROLE IN PLANT ANTIOXIDANT DEFENSE
 CC AND IS POSSIBLY INVOLVED IN NAD(P)/NAD(P)H HOMEOSTASIS.
 CC -1- SIMILARITY: BELONGS TO THE L4BD FAMILY OF NADP-DEPENDENT
 CC OXIDOREDUCTASE.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; Z49268; CAA8262.1; --
 DR EMBL; AL391141; CAC01712.1; --
 DR EMBL; AY091320; AAM14259.1; --
 DR EMBL; AY065253; AAL38729.1; --
 DR PIR; S57612; S57612.
 DR PIR; T51554; T51554.
 DR InterPro; IPR002085; Adh_zn_family.
 DR Pfam; PF00107; Adh_zn_N; 1.
 KW Oxidoreductase; NADP.
 FT CONFLICT 15 15 MISSING (IN REF. 1).
 SQ SEQUENCE 343 AA; 37989 MW; 491698EF58BA82DA CRC64;
 Query Match 7.6%; Score 146; DB 1; Length 343;
 Best Local Similarity 21.6%; Pred. No. 0.0023;
 Matches 86; Conservative 61; Mismatches 160; Indels 92; Gaps 18;
 QY 5 TTOKAVII-----EGDKAVVKTDS--VPELKEGTALVK-----VEAVAGNP 44
 DB 2 TTNKQVLPKHVSQFPKESDFNTTTVELRVPEGSK-SVLVKNLYLSCDPYWRSMGKP 60
 QY 45 TDWKHIAHYKIGPEGLGCDIAGTVKVLGPNASTDLKVGDTGFGFVHGASQTDPKNGAFA 104
 DB 61 DPSSALAAQAPAGPKPIYGYGVSRVIESGHP-----DYKKGLLWGV-----GWE 105

105 EYARVVPPLFYKSNLTHSTADEISEGPKVNFESAASLPVSLTTAGVSLCHLGSKMEWHP 164
 106 EYSVITPMAMHFKIOH-----TDVPLSYITGLMGPGMTAYAGF-----YEV 148
 165 STPOQTHPLLWGGATAVGQOQIQAHHINAYTKIVTVASKKHEKLLKS-YGADDVFDY- 222
 149 CSPKKGTEVTVSAASGAGQVLQGPQAKMGCVV-VGSAGSKEKVDLLTKFGFDDAFNYK 207
 223 HDAGVTEQIKSKYP-NLQHVDAVSGESDIPYAKVTADSLPATLLEVVMTIESIPEEI 281
 208 EESDLSAALKRCRCPKGDIMYFENVGK-----MLDAVLNMMPHGRIAVCGMI 255
 282 RDNVKNIDITLYRASGQ-----ILLGATRFP--ASPEYHEATVKVKFINPHLNG 332
 256 SQ-----YNLENQBGVNLNIIYKRIQGVVADFYDKPKLELVLPRIKEG 305
 333 DIHMMNKFVNSGLDVPALTEGIRKGN--KNVKYVAR 369
 306 KITV--VEDVADGLEKAPALVGLFHGKNGVQVVVIAR 342

RESULT 19
 ADH1_KLULA
 ID ADH1_KLULA STANDARD; PRT; 350 AA.
 AC P20369;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Alcohol dehydrogenase I (EC 1.1.1.1).
 GN ADH1.
 OS Kluyveromyces lactis (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
 OX NCBI_TaxID:28985;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90273773; PubMed=2190430;
 RA Saliola M., Shuster J.R., Falcone C.;
 RT "The alcohol dehydrogenase system in the yeast, Kluyveromyces
 RT lactis.";
 RL Yeast 6:193-204(1990).
 CC -1- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +
 CC NADH.
 CC -1- COFACTOR: Binds 2 zinc ions per subunit.
 CC -1- SUBUNIT: Homotetramer.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
 CC family.
 DR PIR; S09475; S09475.
 DR InterPro; IPR002328; Adh_zinc.
 DR InterPro; IPR002085; Adh_zn_family.
 DR Pfam; PF00107; Adh_zinc_N; 1.
 DR PROSITE; PS00059; ADH_ZINC; 1.
 KW Oxidoreductase; Zinc; Metal-binding; NAD; Multigene family.
 FT METAL 46 46 ZINC 1 (CATALYTIC).
 FT METAL 69 69 ZINC 1 (CATALYTIC).
 FT METAL 100 100 ZINC 2.
 FT METAL 103 103 ZINC 2.
 FT METAL 106 106 ZINC 2.
 FT METAL 114 114 ZINC 2.
 FT METAL 156 156 ZINC 1 (CATALYTIC).
 SQ SEQUENCE 350 AA; 37260 MW; 3D71BE2D5CC86119 CRC64;
 Query Match 7.6%; Score 145.5; DB 1; Length 350;
 Best Local Similarity 23.0%; Pred. No. 0.0025;
 Matches 91; Conservative 49; Mismatches 146; Indels 109; Gaps 18;
 QY 2 SVPTTOKAVII-EGDKAVVKTDSVPELKEGTALVKVEAVAGNP--WK---HAYKI 54
 DB 4 SIPETQKQVIFYENGGELOQKDPVPKRANELLINVKYSGVCHTDLHAWKGDWPLPTKL 63
 QY 55 ----GPEGSILGCDIAGTVKVLGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAEYARV 109


```
Db 64 PLVGGHEG-----AGVVVANGENVK-GWKIGDFAGIKWLGSCMS-----CEYCEL 108
Qy 110 YP-----PLFYKSNLTH-----STADEISBGPVKNFESASLPVS-----LTTAGV 150
Db 109 SNESNCPEADLSGYTHDGSFOQYATADAV-----QAQKIPVGTDLAEVAPVLCAGV 159
Qy 151 SLCHHLSKSM-----EWHPSTPQHTPLLMWGATAVGQOLIQVAKHINAYTKIVTVASKK 206
Db 160 TVYKALKSANLKGADW-----VAISGAAGGLGSLAVQYAKAM-GYRVIGIDAGEE 208
Qy 207 HEKLLKSYGADDVFDYHDAGVI--EQIKSKYPNLOHVLDVAGSEDSIPEAYKVTADSLPA 264
Db 209 KAKLFDKGLGGYFIDFTKSKNIPBEVIEATKGGAGHGVNVSSEFAIEQS-----258
Qy 265 TLEVVWPMPTIESPEEIRKDNVKIDITLLYRASGOEILLGATR-FPASPEYHEATVVKV 323
Db 259 -----TNYVRSNGTVVLVGLPRDAKCKSDVFNQVKSIS 292
Qy 324 FINPLANGDIHNMNIKVFSNGLDDVPALTEGIKE 358
Db 293 IVGSYVGNRADTREAIDPFSSRLVKAPIHVVLGSE 327

RESULT 20
ADH1_CANAL STANDARD; PRT; 350 AA.
AC P43067;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Alcohol dehydrogenase 1 (EC 1.1.1.1) (40 kDa allergen) (Allergen
DE Can a 1) (Can a 1).
GN ADH1 OR CAD.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Microsporid Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=96287648; PubMed=8686375;
RA Betram G., Swoboda R.K., Gooday G.W., Gow N.A.R., Brown A.J.P.;
RT "Structure and regulation of the Candida albicans ADH1 gene encoding
RT an immunogenic alcohol dehydrogenase.";
RL Yeast 12:115-127(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=B311A;
RA Pendrak M.L., Klotz S.A., Smith R.L.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 104-313 FROM N.A.
RA MEDLINE=92136159; PubMed=177830;
RA Shen H.D., Choo K.B., Lee H.H., Heieh J.C., Lin W.L., Lee W.R.,
RA Han S.H.;
RT "The 40-kilodalton allergen of Candida albicans is an alcohol
RT dehydrogenase: molecular cloning and immunological analysis using
RT monoclonal antibodies.";
RL Clin. Exp. Allergy 21:675-681(1991).
CC -!- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +
CC NADH.
CC -!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X81694; CAA57342.1; -.
DR EMBL; U15924; AAA53300.1; -.
DR PIR; S63781; S52153.
DR COMPLEYEST-2DPAGE; P43067; -.
DR InterPro; IPR002328; ADH_zinc.
DR Pfam; PF00107; ADH_zinc_N; 1.
DR PROSITE; PS00059; ADH_ZINC; 1.
KW Oxidoreductase; Zinc; Metal-binding; NAD; Allergen.
FT METAL 46 46 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 69 69 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 100 100 ZINC 2 (BY SIMILARITY).
FT METAL 103 103 ZINC 2 (BY SIMILARITY).
FT METAL 106 106 ZINC 2 (BY SIMILARITY).
FT METAL 114 114 ZINC 2 (BY SIMILARITY).
FT METAL 156 156 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT CONFLICT 39 39 H -> N (IN REF. 2).
FT CONFLICT 53 53 R -> W (IN REF. 2).
FT CONFLICT 140 140 A -> T (IN REF. 3).
FT CONFLICT 212 212 P -> L (IN REF. 2).
FT CONFLICT 227 227 D -> A (IN REF. 3).
FT CONFLICT 313 313 R -> S (IN REF. 3).
FT CONFLICT 327 327 D -> E (IN REF. 2).
SQ SEQUENCE 350 AA; 36879 MW; C3DPE2E70F42D634 CRC64;

Query Match 7.5%; Score 143.5; DB 1; Length 350;
Best Local Similarity 22.4%; Pred. No. 0.0035;
Matches 94; Conservative 57; Mismatches 123; Indels 145; Gaps 24;

Qy 3 VPTQKAVIIE--GDKAVKTDVSVPELKEGTALVKVE-----AVAGNPTDWHIAYK 53
Db 5 IPKTQKAVVDTNGGQLVYK-DYVPTPKPKNELLHVYKSGVCHTDLHARKGDWP-LATK 62
Qy 54 I----GPEGILGCDIAGTVVKGPNASTDLKVG-D-TGFGFVHGASOT-----DPK- 99
Db 63 LPLVGGHEG-----AGVVVGMGENVK-GWKIGDFAGIKWLGSCMSCBPQCGABPNC 114
Qy 100 -----NGAPAEYARVYVPLPYKSNLTHSTADEISBGPVKNFESASLPVSLTTAG 149
Db 115 GEADLSGYTHDGSFEQYA-----TADAV-----QAQKIPAGTOLAN 150
Qy 150 VS--LCHHLG-----SKMEWHPSTPQHTPLLMWGATAVGQOLIQVAKHINAYT 197
Db 151 VAPILCAGTVYKALKTADLAAGW-----VAISGAAGGLGSLAVQYARAMGLRV 200
Qy 198 KIVTVASKKHEKLLKSYGADDVFDYHDAGVIEQIKSKYPNLOHVLDVAGSEDSIPEAYKV 257
Db 201 VAIDGGDEKGE-FVKSIGAEAYVDF-----TKDKDIVEAVKK 236
Qy 258 TADSLPATLLEVVWPMPTIESPEEIRKDNVKIDITLLY-RASGOEILLGATRPPA-----S 311
Db 237 ATDGGPHGANV-----SVSEK-----AIDQSVYVRPLGKVLVVG--LPAHAKVTA 281
Qy 312 PEYHEATVVKVFPINPLANGDIHNMNIKVFSN-----GLDDVPALTEGIKEGK 360
Db 282 PVP-DAVVKSIIRKSGSVGNRKDTAEALDFPSRLCIKPIKIVGLSGLPEVFKLMBSGK 339

RESULT 21
ID_TDH_PYRHO STANDARD; PRT; 348 AA.
AC O58389;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable L-threonine 3-dehydrogenase (EC 1.1.1.103).
GN TDH OR PH0655.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
```


QY 106 YARVPPFLFYKSNLTHSTADISE--GPVKNFESASLPVSLTTAGVSLCHHLSKMEWH 163
Db 124 YLVI--PAFNAFKIPDNISDOLAFDP---FGNAVHTALSFDLVGD-----166
QY 164 PSTQHTPLLIWGATAVGOOLIQVAKHINAYTKIVTVASKHEKLLKSYGADDDVFDY- 222
Db 167 -----VLVSGAGPIGIMAAVAKHVGARNVIT-----DVNEYR 200
QY 223 ----HDAGVIEQIKSKYPNLQHVDAVGSBSI-----PEAYKVTADSL-----PAT 265
Db 201 LELARKMGITRAVNVAKENLNDVMTLQWTEGDFVGLSEMGAPPAFTMLDTMNHGGRIA 260
QY 266 LLEVPMWITIEPIEIRKONKVIDIT-LLYRASQOEILLGLTRPPASPEYHEATVKVKF 324
Db 261 MLGIPP-----SDMSIDWTKVIFKGLFTKIGYREMP-----ETWYKMAAL 301
QY 325 INPHLNGDI--HNMNIKVFSGLDVDPALTEGIEK 360
Db 302 IQSGLDLSPILTHRSIDDFQKGFDP---AMRSG-QSGK 335
RESULT 23
ADH1_BACST STANDARD; PRT; 337 AA.
AC Fl311;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Alcohol dehydrogenase [EC 1.1.1.1] (ADH-T).
GN ADH1.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.
OX NCBI TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A., AND MUTAGENESIS.
RC STRAIN=NCA 1503;
RX MEDLINE=92138636; PubMed=1735726;
RA Sakoda H., Inanaka T.;
RT "Cloning and sequencing of the gene coding for alcohol dehydrogenase
of Bacillus stearothermophilus and rational shift of the optimum
pH";
RT J. Bacteriol. 174:1397-1402(1992).
RN [2]
RP SEQUENCE OF 1-45.
RX MEDLINE=7329257; PubMed=4578954;
RA Bridgen J., Kolb E., Harris J.I.;
RT "Amino acid sequence homology in alcohol dehydrogenase";
RL FEBS Lett. 33:1-3(1973).
RN [3]
RP SEQUENCE OF 34-54.
RX MEDLINE=79169263; PubMed=436831;
RA Jeck R., Woenckhaus C., Harris J.I., Runswick M.J.;
RT "Identification of the amino acid residue modified in Bacillus
stearothermophilus alcohol dehydrogenase by the NAD+ analogue 4-(3-
bromoacetylpyridinio)butyldiphosphoadenosine";
RL Eur. J. Biochem. 93:57-64(1979).
RN [4]
RP SEQUENCE OF 1-37; 188-197; 247-263 AND 324-336.
RC STRAIN=NCA 1503;
RX MEDLINE=94325354; PubMed=8049268;
RA Robinson G.A., Bailey C.J., Dowds B.C.A.;
RT "Gene structure and amino acid sequences of alcohol dehydrogenases of
Bacillus stearothermophilus";
RL Biochim. Biophys. Acta 1218:432-434(1994).
CC -1- FUNCTION: THERMOSTABLE NAD(+) DEPENDENT ALCOHOL DEHYDROGENASE.
CC -1- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +
NADH.
CC -1- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
CC -1- ENZYME REGULATION: SUBSTRATE INHIBITION IS NOT OBSERVED WITH ANY
ALCOHOLS. AND THE ENZYME-NADH DISSOCIATION IS NOT CONSIDERED TO BE
A RATE-LIMITING STEP.
CC -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL: D90421; BAA14411.1; -;
DR PIR: A42654; A42654.
DR InterPro: IPR002328; ADH_zinc.
DR InterPro: IPR002085; Adh_zn_family.
DR Pfam: PF0107; ADH_zinc_N; 1.
DR PROSITE: PS0059; ADH_ZINC; 1.
KW Oxidoreductase; Zinc; Metal-binding; NAD.
FT METAL 38 38 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 61 61 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 92 92 ZINC 2 (BY SIMILARITY).
FT METAL 95 95 ZINC 2 (BY SIMILARITY).
FT METAL 98 98 ZINC 2 (BY SIMILARITY).
FT METAL 106 106 ZINC 2 (BY SIMILARITY).
FT METAL 148 148 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT MUTAGEN 40 40 T->S: LITTLE DECREASE IN ACTIVITY.
FT MUTAGEN 43 43 H->R: HIGHER LEVEL OF ACTIVITY AT PH 9.
FT MUTAGEN 38 38 C->S: NO ACTIVITY.
FT MUTAGEN 40 40 T->A: NO ACTIVITY.
FT MUTAGEN 43 43 H->A: NO ACTIVITY.
FT CONFLICT 22 22 MISSING (IN REF. 2).
FT CONFLICT 33 33 MISSING (IN REF. 2).
FT CONFLICT 52 53 KP -> PK (IN REF. 3).
SQ SEQUENCE 337 AA; 36100 MW; B9B35A80E9B7A86 CRC64;
Query Match 7.4%; Score 141.5; DB 1; Length 337;
Best Local Similarity 24.5%; Pred. No. 0.0047;
Matches 90; Conservative 57; Mismatches 152; Indels 69; Gaps 18;
QY 8 KAVIIEGDKAVVKT-DVSVPELKEGTALVKVEAVAGNPTDW--KHAYKIGPEGSIL-GC 63
Db 2 KAAVVEQFKPLQVKEVEKPKISYGEVLVRIKACGVCHTDLHAAHGDPVKPLPLPGH 61
QY 64 DIAGTVVKLGPNASTDLKVG-DTGFVHGA-----SQTDPKNGAF 103
Db 62 EGVGVIEEVGPGV--THLKVDVRGVPWLYSACGCHDYCLSGQETLCERQGNAGVSDGGY 120
QY 104 AEYAVYVPLFYKSNLTHSTADISEGPVK-NFESASLPVSLTTAGVSLCHL---GSK 159
Db 121 ASYCR-----AAADYVVKIPDNLSFEAA-----PIPCAGVTTYKALKVTGAK 163
QY 160 M-EWHPSTPQHTPLLIWGATAVGOOLIQVAKHINAYTKIVTVASKHEKLLKSYGADD 218
Db 164 PGW-----VAIY-GIGGLGHVAVQYAKAMGLNVAVDLGDEKLE-LAKQLGADL 211
QY 219 VFYDHDAGVIEQIKSKYPNLQHVDAVGSBSIPEAYKVTADSLPATLLEWPMWITIEIP 278
Db 212 VVNPKHDDAAQWIKKVGGVHATVAVSKAAPESAYKIRRGACVGLVGLP---BEIP 268
QY 279 BEIRKONKVIDITLLYRASG-QEILLGATRFPPASPEYHEATVKVKFINPHLNGDIHMH 337
Db 269 IPT-FDTVLNGVKIIGSIVGTRKDLQALQFAA-----EGKVKITVEVPLENINDVDR 322
QY 338 NIKVFSNG 345
Db 323 MLKGQING 330
RESULT 24
TDH_PYRAB
ID TDH_PYRAB STANDARD; PRT; 348 AA.
AC Q9UYX0;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)


```
FT METAL          99      ZINC 2 (BY SIMILARITY).
FT METAL         107      ZINC 2 (BY SIMILARITY).
FT METAL         148      ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT CONFLICT       54      T -> N (IN REF. 2).
SQ SEQUENCE      341 AA; 37202 MW; AAB2877DD5FAC1E CRC64;

Query Match          7.3%; Score 141; DB 1; Length 341;
Best Local Similarity 20.4%; Pred. No. 0.0052;
Matches 81; Conservative 52; Mismatches 147; Indels 118; Gaps 18;

QY 12 IEQDKAVVKTDSVPPELKEGTALVKVEAVAGNPTD-----WKHAYKIGPEGSILGCDIA 66
DB 7 LKAEIGIWMTDVPPELGHNDLLIKIRKTAICGTDVHIYNWDEWSQKTIPTVPMVVGHEYV 66
QY 67 GTVVKLGPNASTDLKVGDTGFGFVH-----GASQTDPRKNGAF 105
DB 67 GEVVGIGQEVK-GPKIGDRVSGEGHITCGHRCRCGRTHLCRNTIGVGNRP--GCFAE 123
QY 106 YARVYPLFYKSNLTHSTADEISE--GPVKNFESAASLPVSLTTAGVSLCHLGSKMEWH 163
DB 124 YLVI--PAPNAFKIPDNISDLSIFDP--FGNAVHTALSFDLVGED-----166
QY 164 PSTPQHTPLLIWGATAVGQOLIQAQKHINAYTKIVTASKEKHLKLSYGADDDVDPY- 222
DB 167 -----VLVSGAGPIGIMAAVAKHVGARNVIT-----DVNEYR 200
QY 223 -----HDAGVIEQIKSKYPNLQHVDAVGSBDSI-----PEAYKVTADSL-----PAT 265
DB 201 LELARKNGITRAVNVAKENLNDVMAELGTEGPDVGLMSGAPPARTMLDTMNHGRIA 260
QY 266 LLEVVPMTIESIPBEIRKONVKIDIT--LLYRASQEIILLGATRPASPPEYHEATVKPKV 324
DB 261 MLGIPP-----SDMSIDWTKVIFKGLFIKGIYGREMF-----ETWYKMAAL 301
QY 325 INPHLNGDI--HHMNIKVFNSGLDDVPALTEGKEGK 360
DB 302 IQSGLDLSPIITHRFSIDDFQKGFDP-----AMCSG-QSGK 335

RESULT 26
TDH_SHIFL
ID TDH_SHIFL STANDARD; PRT; 341 AA.
AC PS9409;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE L-threonine 3-dehydrogenase [EC 1.1.1.103].
GN TDH OR SF3656.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=2272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RL through comparison with genomes of Escherichia coli K12 and O157."
RL Nucleic Acids Res. 30:4432-4441(2002).
CC -1- CATALYTIC ACTIVITY: L-threonine + NAD(+) = L-2-amino-3-
CC oxobutanoate + NADH.
CC -1- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
CC -1- PATHWAY: Threonine catabolism; first step.
CC -1- SUBUNIT: Homotetramer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
CC family.
-----
```

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1eb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AE015373; AAN45103.1; -.
DR HAMAP; MF 00627; -.
DR PROSITE; PS00059; ADH_ZINC; 1.
KW Oxidoreductase; NAD; Zinc; Metal-binding.
FT METAL 38 38 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 63 63 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 93 93 ZINC 2 (BY SIMILARITY).
FT METAL 96 96 ZINC 2 (BY SIMILARITY).
FT METAL 99 99 ZINC 2 (BY SIMILARITY).
FT METAL 107 107 ZINC 2 (BY SIMILARITY).
FT METAL 148 148 ZINC 1 (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 341 AA; 37255 MW; AAB2877DC84AAC1E CRC64;

Query Match          7.3%; Score 141; DB 1; Length 341;
Best Local Similarity 20.4%; Pred. No. 0.0052;
Matches 81; Conservative 52; Mismatches 147; Indels 118; Gaps 18;

QY 12 IEQDKAVVKTDSVPPELKEGTALVKVEAVAGNPTD-----WKHAYKIGPEGSILGCDIA 66
DB 7 LKAEIGIWMTDVPPELGHNDLLIKIRKTAICGTDVHIYNWDEWSQKTIPTVPMVVGHEYV 66
QY 67 GTVVKLGPNASTDLKVGDTGFGFVH-----GASQTDPRKNGAF 105
DB 67 GEVVGIGQEVK-GPKIGDRVSGEGHITCGHRCRCGRTHLCRNTIGVGNRP--GCFAE 123
QY 106 YARVYPLFYKSNLTHSTADEISE--GPVKNFESAASLPVSLTTAGVSLCHLGSKMEWH 163
DB 124 YLVI--PAPNAFKIPDNISDLSIFDP--FGNAVHTALSFDLVGED-----166
QY 164 PSTPQHTPLLIWGATAVGQOLIQAQKHINAYTKIVTASKEKHLKLSYGADDDVDPY- 222
DB 167 -----VLVSGAGPIGIMAAVAKHVGARNVIT-----DVNEYR 200
QY 223 -----HDAGVIEQIKSKYPNLQHVDAVGSBDSI-----PEAYKVTADSL-----PAT 265
DB 201 LELARKNGITRAVNVAKENLNDVMAELGTEGPDVGLMSGAPPARTMLDTMNHGRIA 260
QY 266 LLEVVPMTIESIPBEIRKONVKIDIT--LLYRASQEIILLGATRPASPPEYHEATVKPKV 324
DB 261 MLGIPP-----SDMSIDWTKVIFKGLFIKGIYGREMF-----ETWYKMAAL 301
QY 325 INPHLNGDI--HHMNIKVFNSGLDDVPALTEGKEGK 360
DB 302 IQSGLDLSPIITHRFSIDDFQKGFDP-----AMRSG-QSGK 335

RESULT 27
ADH2_KLUMA
ID ADH2_KLUMA STANDARD; PRT; 347 AA.
AC QSP4C2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Alcohol dehydrogenase II (EC 1.1.1.1).
GN ADH2.
OS Kluyveromyces marxianus (Yeast) (Kluyveromyces fragilis).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=4911;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12424;
RX MEDLINE=20432119; PubMed=10974568;
RA Ladriere J.-M., Georis I., Guerin M., Vandenhaute J.;
RT "Kluyveromyces marxianus exhibits an ancestral Saccharomyces
-----
```

```
RT cerevisiae genome organization downstream of ADH2.1;
RL Gene 255:83-91(2000).
CC -1- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +
CC NADH.
CC -1- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
CC -1- SUBUNIT: Homotrimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF225206; AAF91235.1; -.
DR InterPro; IPR002328; ADH_zinc.
DR Pfam; PF00107; ADH_zinc_N; 1.
DR PROSITE; PS00059; ADH_ZINC; 1.
DR Oxidoreductase; Zinc; Metal-binding; NAD; Multigene family;
KW Acetylation.
FT INIT_MET 0 0 BY SIMILARITY.
FT MOD_RES 43 1 ACETYLATION (BY SIMILARITY).
FT METAL 43 43 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 66 66 ZINC 2 (CATALYTIC) (BY SIMILARITY).
FT METAL 97 97 ZINC 2 (BY SIMILARITY).
FT METAL 100 100 ZINC 2 (BY SIMILARITY).
FT METAL 103 103 ZINC 2 (BY SIMILARITY).
FT METAL 111 111 ZINC 2 (BY SIMILARITY).
FT METAL 153 153 ZINC 1 (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 347 AA; 36837 MW; F70DD4034F0CB49 CRC64;
Query Match 7.3%; Score 140.5; DB 1; Length 347;
Best Local Similarity 21.5%; Pred. No. 0.0058;
Matches 89; Conservative 54; Mismatches 138; Indels 133; Gaps 20;
QY 2 SVPTQKAVII-EGDKAVVKTDSVPBELKGTALVKVEAVAGNPTD---WK-----HI 50
Db 1 SIPTTQKGVIFYENGQQLYKDIIPVPRKPSNELLINVKYSGVCHTDLHAWKGDWPLDTKL 60
QY 51 AYKIGPGSILGCDIAGTVVVLGNASTDLKVED-TGFGFVHGASQTDPK-----99
Db 61 PLVGGHEG-----AGVWVAMGDNVK-GWKIGDLGKWLNGSCMNCCELSNESNCP 112
QY 100 -----NGAFAYEARVYPPFLFYKSNLTHSTADEISEGPKVKNFESAASLPVS-----144
Db 113 DADLSGYTHDGSFOQYA-----TDAV-----QAHIIPAGTDLAQV 148
QY 145 --LFTAGVSLCHILGSKM-----EWHPSTPQHTPELLIWGGATAVGQOLIQVAKHINAYTK 198
Db 149 APILCAGVTYVYKALKTAEMKAGDW-----VAISGAAGGLGSLAVQYAKAM-GFRV 197
QY 199 IVTVASKKHEKLKSYGADVDVFDHAGVIEQIKSKYPNLQHVLDVAGSSEDSIPEAYKVT 258
Db 198 LGIDGGEGKEBELFKSLGG-EVF-----IDFTKSKDIQVGEVIKAT 235
QY 259 ADSLPAFL-LEVPWMTTETIEPIEIRKDNVKIDITLLYRASGOETLLGATR-PPASPEYHE 316
Db 236 NGGAHVGVINSVSEKALESIE-----YCRSNGTVVLVGLPKDAKCKSDVFN 282
QY 317 ATVKFVKFIPHL-NGGDIH-----MNKVFNSGLDDVPALTEGIEKEGK 360
Db 283 QVVKSIHIVSGYVGNRAUTREALDFCKGLGNAPIKVVGSLTLPPIEYKEMEQKG 336
RESULT 28
TDH_ECOLI
ID -TDH_ECOLI
AC P07913;
```

```
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE L-threonine 3-dehydrogenase (EC 1.1.1.103).
GN TDH OR B3616.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-20 AND 149-153.
RC STRAIN=K12;
RX MEDLINE=89174812; PubMed=2647748;
RA Aronson B.D., Somerville R.L., Epperly B.R., Dekker E.E.;
RT "The primary structure of Escherichia coli L-threonine
RT dehydrogenase.";
RL J. Biol. Chem. 264:5226-5232 (1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=94316500; PubMed=8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes.";
RL Nucleic Acids Res. 22:2576-2586 (1994).
RN [3]
RP ACTIVE SITE, REQUIREMENT FOR ZINC, AND PARTIAL SEQUENCE.
RX MEDLINE=91177850; PubMed=2007567;
RA Epperly B.R., Dekker E.E.;
RT "L-threonine dehydrogenase from Escherichia coli. Identification of
RT an active site cysteine residue and metal ion studies.";
RL J. Biol. Chem. 266:6086-6092 (1991).
CC -1- CATALYTIC ACTIVITY: L-threonine + NAD(+) = L-2-amino-3-
CC oxobutanoate + NADH.
CC -1- COFACTOR: Binds 2 zinc ions per subunit. Cobalt, cadmium and
CC manganese can exchange for zinc.
CC -1- ENZYME REGULATION: Activated by manganese or cobalt ions.
CC -1- PATHWAY: Threonine catabolism; first step.
CC -1- SUBUNIT: Homotrimer.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X06690; CAA29884.1; -.
DR EMBL; U00039; AAB18593.1; -.
DR EMBL; AE000439; AAC76640.1; -.
DR PIR; A33276; DEECTH.
DR ECO2DBASE; G038.8; 6TH EDITION.
DR EcoGene; EG10993; tdh.
DR HAMAP; MF_00627; -.
DR InterPro; IPR002328; ADH_zinc.
DR InterPro; IPR002085; Adh_zn_family.
DR InterPro; IPR000205; NAD_binding.
DR InterPro; IPR004627; Tdh.
DR Pfam; PF00107; ADH_zinc_N; 1.
DR TIGRFAMs; TIGR00692; tdh; 1.
DR PROSITE; PS00059; ADH_ZINC; 1.
DR Oxidoreductase; Zinc; Metal-binding; Manganese; Cobalt; NAD;
KW Complete proteome.
FT METAL 38 38 ZINC 1 (CATALYTIC).
FT METAL 63 63 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 93 93 ZINC 2 (BY SIMILARITY).
FT METAL 96 96 ZINC 2 (BY SIMILARITY).
FT METAL 99 99 ZINC 2 (BY SIMILARITY).
FT METAL 107 107 ZINC 2 (BY SIMILARITY).
```


FT METAL 148 148 ZINC 1 (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 341 AA; 37239 MW; 039FBD6B1CE8C2B2 CRC64;

Query Match
Best Local Similarity 7.3%; Score 140; DB 1; Length 341;
Matches . 80; Conservative 52; Mismatches 150; Indels 114; Gaps 17;

Qy 12 IEGDKAVKVTDSVPELKEGTALVKEAVAGNPTD-----WKHIAYKIGPEGSILGCDIA 66
Db 7 LKAEIGIWMTDVPPVPELGHNDLLIKIRKTAICGTDVHIYNWDEWSQKTIPIVPMVVGHEYV 66

Qy 67 GTVVKLGPNASTDLKVGDTGFGFVH-----GASQTDPPKNGAPAE 105
Db 67 GEVVDIGQEVK-GFKIGDRVSGEGHITCGHCRNCRGRTLCRNTIGVGNRP--GCPAE 123

Qy 106 YARVPPPLFYKSNLTHSPADISSEGVKVPKPSAASLPSVLTAGVSLCHLGSKMEWHP 165
Db 124 YLVI--PAFNAFKPDNISDDIA-AIFPFGNAVHTALSFLVGED----- 166

Qy 166 TPQHTHPLLLINGGATAVGOQLIQVAKHINAYTKIVTVASKKHEKLLKSYGADDDVFDY--- 222
Db 167 -----VLVSGAGPIGIMAAVAKHVGARNVIT-----DVNEYRLE 202

Qy 223 --HDAGVTEQIKSKYKPNLQHVDAVGSDSI-----PEAVKVTADSL-----PATLL 267
Db 203 LARKONGITRAVNVAKENLNDVMAELGTEGFDVGLMSGAPPAPFRMTLDTMNHGGRIAML 262

Qy 268 EVVPMWTIESIPEIRKONKVKIDIT--LLVRASQEIILGATRPASPSPHEATVFKVFKN 326
Db 263 GIPP-----SDMSIDWTKVIFKGIYGREMP-----ETWYKQALIQ 303

Qy 327 PHLANGDI--HHMNIKVSNGLDVPALEGIKEGK 360
Db 304 SGLDLSPIITHRESIDDFQKGFDP---AMRSG-QSGK 335

RESULT 29
ID _TDH SHEON STANDARD; PRT; 341 AA.
AC Q8E8J1,
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE L-threonine 3-dehydrogenase (EC 1.1.1.103).
GN TDH OR S04673.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1,
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vanathevan J., Weidman J., Imprim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA Feilblyum T.V., Smith H.O., Venter J.C., Nealeon K.H., Fraser C.M.,
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
Shewanella oneidensis".
RL Nat. Biotechnol. 20:1118-1123(2002).
CC -1- CATALYTIC ACTIVITY: L-Chreonine + NAD(+) = L-2-amino-3-
CC oxobutanate + NADH.
CC -1- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
CC -1- PATHWAY: Threonine catabolism; first step.
CC -1- SUBUNIT: Homotetramer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.ebi.ac.uk/announcements/
or send an email to license@ebi.ac.uk).

EMBL; S04673; AAN57632.1; --
DR TIGR; S04673; --
DR HAMAP; MF_00627; --
DR InterPro; IPR002328; ADH_zinc.
DR InterPro; IPR002085; Adh zn family.
DR InterPro; IPR002025; NAD_binding.
DR InterPro; IPR004627; Tdh.
DR Pfam; PF00107; ADH_zinc_N; 1.
DR TIGRFAMs; TIGR00692; tdn; 1.
DR PROSITE; PS00059; ADH_ZINC; 1.
KW Oxidoreductase; NAD; Zinc; Metal-binding; Complete proteome.
FT METAL 38 38 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 63 63 ZINC 2 (CATALYTIC) (BY SIMILARITY).
FT METAL 93 93 ZINC 2 (BY SIMILARITY).
FT METAL 96 96 ZINC 2 (BY SIMILARITY).
FT METAL 99 99 ZINC 2 (BY SIMILARITY).
FT METAL 107 107 ZINC 2 (BY SIMILARITY).
FT METAL 148 148 ZINC 1 (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 341 AA; 37255 MW; 5378A4501153C7BF CRC64;

Query Match 7.3%; Score 140; DB 1; Length 341;
Best Local Similarity 21.4%; Pred. No. 0.0061;
Matches 81; Conservative 59; Mismatches 138; Indels 100; Gaps 20;

Qy 12 IEGDKAVKVTDSVPELKEGTALVKEAVAGNPTD-----WKHIAYKIGPEGSILGCDIA 66
Db 7 LKAEIGIWMTDVPPVPELGHNDLLIKIRKTAICGTDVHIYNWDEWSQKTIPIVPMVVGHEYV 66

Qy 67 GTVVKLGPNASTDLKVGDTGFGFV-----HGASQTD-----PKNGAPAEYA 107
Db 67 GEVVDIGQEVK-RGFNIGDRVSGEGHITCGHCRNCRGRTLCRNTSGVGNREGSPAEYL 125

Qy 108 RVYPPPLFYKSNLTHSTADEISE--GPVN--FESASL-----PVSLLTAG-----VSLC 153
Db 126 VI--PAFNAFKIPDDISDDLASIPDFGNVHTALSFLVGEDVLIITGAGPIGIMAAVAVC 183

Qy 154 HHLGSKMEWHPSTPQHTHPLLLINGGATAVGOQLIQVAKHINAYTKIVTVASKKHEKLLKS 213
Db 184 RHVGAR-----HVV-----TDVNEYRLEARKMGA-TRAVNVAGENLKVWKE 226

Qy 214 YGADDDVFDYHDAGVIEQIKSKYKPNLQHVDAVGSDES-1PEAYKVTADSLPATLLEVP 272
Db 227 LGWTEGFD-----VGLMSGVPSAPRAMLD-----TMNHGSKI 259

Qy 273 TISIP-EEIRKONKVKIDITLLVRASQEIILGATRPASPSPHEATVFKVPKPNPHLN 331
Db 260 AMLGIPGEMADWSKY-----IFKGLVIGKIYGREMP-----ETWYKMASLQSGLDI 308

Qy 332 GDI--HHMNIKVSNGLD 347
Db 309 SPIITHYKIDDFQKGFDP 326

RESULT 30
ID _VAT1 TORCA STANDARD; PRT; 379 AA.
AC P19333;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Synaptic vesicle membrane protein VAT-1.
OS Torpedo californica (Pacific electric ray).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Squalia; Hypnosqualia; Pristiogalea; Batoida;
OC Torpediniformes; Torpedinoidei; Torpedinidae; Torpedo.

Search completed: January 10, 2004, 10:05:32
Job time : 29 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2004, 09:49:00 ; Search time 84 Seconds
(without alignments)

1136.660 Million cell updates/sec

Title: US-10-081-644-2

Perfect score: 1920

Sequence: 1 MSVPTTKAVIIIEGKAVK.....ALTEGIKEGKNVYVARL 370

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 90 summaries

Database :

SPTREMBL 23.:

- 1: sp_archaea.:
- 2: sp_bacteria.:
- 3: sp_fungi.:
- 4: sp_human.:
- 5: sp_invertebrate.:
- 6: sp_mammal.:
- 7: sp_mhc.:
- 8: sp_organelle.:
- 9: sp_phase.:
- 10: sp_plant.:
- 11: sp_rodent.:
- 12: sp_virus.:
- 13: sp_vertebrate.:
- 14: sp_unclassified.:
- 15: sp_rvirus.:
- 16: sp_bacteriap.:
- 17: sp_archaeap.:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	309.5	16.1	348	3 Q9P7F4	Q9P7F4 schizosacch
2	294	15.3	362	3 Q8U0F9	Q8U0F9 penicillium
3	277.5	14.5	339	10 Q941I0	Q941I0 fragaria an
4	271	14.1	337	10 Q23939	Q23939 fragaria ve
5	266.5	13.9	363	3 Q9Y7D0	Q9Y7D0 aspergillus
6	264	13.8	313	16 Q8Y9B9	Q8Y9B9 listeria mo
7	261	13.6	386	10 Q943P3	Q943P3 arabidopsis
8	260	13.5	313	16 Q92E39	Q92E39 listeria in
9	259	13.5	317	10 Q43677	Q43677 vigna radia
10	248.5	12.9	309	10 Q8H2V0	Q8H2V0 oryza sativ
11	248	12.9	309	10 Q94T67	Q94T67 helianthus
12	247	12.9	309	10 Q8LEB8	Q8LEB8 arabidopsis
13	219	11.4	311	16 Q8ES25	Q8ES25 oceanobacil
14	218	11.4	353	16 Q8UK00	Q8UK00 agrobacteri
15	214	11.1	313	16 Q85702	Q85702 streptomyce
16	208	10.8	324	16 Q9KFW8	Q9KFW8 bacillus ha

17	200	10.4	309	17	Q8TSM3	Q8tam3 methanoearc
18	199.5	10.4	329	10	Q8HOM1	Q8hOm1 epinacia ol
19	193	10.1	343	16	Q67374	Q67374 aquifex ao
20	190	9.9	322	16	Q8DV65	Q8dv65 streptococc
21	189	9.8	308	16	Q8HL3	Q8hl3 rhizobium l
22	188.5	9.8	333	16	Q8FLN0	Q8fln0 corynebacte
23	188	9.8	339	16	Q9L2A5	Q9l2a5 streptomyce
24	187.5	9.8	322	16	Q9KEB8	Q9keb8 bacillus ha
25	187	9.7	359	16	Q8UJY9	Q8ujy9 agrobacteri
26	182	9.5	339	10	Q9AYU1	Q9ayU1 triphyesaria
27	178.5	9.3	335	16	Q8CU46	Q8cu46 staphylococ
28	176	9.2	329	16	Q8RJR7	Q8rjr7 streptomyce
29	176	9.2	331	17	Q97V32	Q97v32 sulfolobus
30	174.5	9.1	358	16	Q8UKD5	Q8ukd5 agrobacteri
31	173	9.0	328	16	Q9CHL2	Q9chl2 lactococcus
32	172.5	9.0	339	16	Q9KEW1	Q9kew1 bacillus ha
33	172.5	9.0	340	16	Q8PC20	Q8pc20 xanthomonas
34	172	9.0	337	16	Q92R74	Q92r74 rhizobium m
35	171.5	8.9	338	16	Q8UGN9	Q8ugn9 agrobacteri
36	170.5	8.9	326	16	Q8FYW9	Q8fyw9 brucella su
37	170.5	8.9	336	16	Q8YIY7	Q8yiy7 brucella me
38	168.5	8.8	244	3	Q8J284	Q8j284 podospora a
39	168.5	8.8	340	16	Q8PNP3	Q8pnp3 xanthomonas
40	168	8.8	334	16	Q8J41	Q8j41 rhizobium l
41	168	8.8	337	16	Q8NLD8	Q8nid8 corynebacte
42	167.5	8.7	288	10	Q9SKF6	Q9skf6 lithospermu
43	167	8.7	366	10	Q8LBA4	Q8lba4 arabidopsis
44	167	8.7	366	10	Q9LK96	Q9lk96 arabidopsis
45	165	8.6	332	17	Q96YB5	Q96yb5 sulfolobus
46	164.5	8.6	329	10	Q9SV68	Q9sv68 arabidopsis
47	164.5	8.6	348	16	Q8UH60	Q8uh60 agrobacteri
48	164	8.5	337	16	Q9KUG9	Q9kug9 vibrio chol
49	163	8.5	335	5	Q8T122	Q8t122 dictyosteli
50	161.5	8.4	335	16	Q53146	Q53146 mycobacteri
51	159.5	8.3	348	16	Q8YZP5	Q8yzp5 anabaena sp
52	158.5	8.3	354	4	Q9BRA4	Q9bra4 homo sapien
53	158.5	8.3	363	16	Q985B6	Q985b6 rhizobium l
54	158	8.2	326	16	Q98F10	Q98f10 rhizobium l
55	158	8.2	337	16	Q8EJR1	Q8ejr1 shewanella
56	157	8.2	328	5	Q45496	Q45496 caenorhabdi
57	157	8.2	360	10	Q8RUH4	Q8ruh4 oryza sativ
58	156.5	8.2	325	16	Q8PP94	Q8pp94 xanthomonas
59	156.5	8.2	8563	2	Q54297	Q54297 streptomyce
60	156.5	8.2	10223	2	Q54296	Q54296 streptomyce
61	155	8.1	328	16	P72043	P72043 mycobacteri
62	155	8.1	335	16	Q8P362	Q8p362 xanthomonas
63	154.5	8.0	316	16	Q8DC41	Q8dc41 vibrio vuln
64	154.5	8.0	2504	16	Q92851	Q92851 rhizobium m
65	153.5	8.0	326	16	Q9XA55	Q9xa55 streptomyce
66	153.5	8.0	336	3	Q42703	Q42703 candida par
67	153	8.0	337	16	Q8DMW0	Q8dmw0 synchococc
68	152	7.9	336	16	Q8ZFP1	Q8zfp1 yersinia pe
69	152	7.9	348	16	Q8DOP7	Q8dop7 yersinia pe
70	151	7.9	325	3	Q74822	Q74822 schizosacch
71	151	7.9	325	10	Q65423	Q65423 arabidopsis
72	151	7.9	335	16	Q8PEK0	Q8pek0 xanthomonas
73	150.5	7.8	396	4	Q8MWV3	Q8mwv3 homo sapien
74	150	7.8	331	17	Q8ZUP0	Q8zup0 pyrobaculum
75	150	7.8	13414	5	Q81616	Q81616 cryptospori
76	149.5	7.8	322	4	Q14679	Q14679 homo sapien
77	149.5	7.8	332	4	Q9BWB8	Q9bwb8 homo sapien
78	149.5	7.8	388	16	Q9RVG8	Q9rvg8 deinooccus
79	149	7.8	318	16	Q8NTU9	Q8ntu9 corynebacte
80	148.5	7.7	336	16	Q8CKJ7	Q8ckj7 staphylococ
81	148.5	7.7	348	10	Q9LX24	Q9lxx4 arabidopsis
82	147.5	7.7	351	10	Q8LA26	Q8la26 arabidopsis
83	147	7.7	328	16	Q914J8	Q914j8 pseudomonas
84	147	7.7	340	5	Q25222	Q25222 leishmania
85	147	7.7	343	10	Q8LD14	Q8ld14 arabidopsis
86	147	7.7	447	16	Q9ZBK1	Q9zbk1 streptomyce
87	146.5	7.6	338	17	Q97WA1	Q97wa1 sulfolobus
88	146.5	7.6	351	10	Q9C677	Q9c677 arabidopsis
89	146.5	7.6	396	4	Q96RL8	Q96rl8 homo sapien

```
90 146.5 7.6 484 13 Q8JFV8 08Jfv8 brachydanio
ALIGNMENTS
RESULT 1
Q9P7F4 PRELIMINARY; PRT; 348 AA.
AC Q9P7F4
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative dehydrogenase by similarity.
GN SPAC2E1P3.01.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972h-;
RA Harris D., Wood V., Rajandream M.A., Barrell B.G.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL162531; CAB83005.1; -
DR GenDB SPombe; SPAC2E1P3.01; -
DR InterPro; IPR002085; Adh_zn_family.
DR Pfam; PF00107; adh_zinc; 1.
SQ SEQUENCE 348 AA; 37929 MW; 63E638A712AFF1C2 CRC64;
Query Match 16.1%; Score 309.5; DB 3; Length 348;
Best Local Similarity 28.2%; Pred. No. 1.7e-15;
Matches 110; Conservative 49; Mismatches 160; Indels 71; Gaps 13;
Qy 8 KAVIIIEGDKAV-VKTDVSVPELKEGTALVKVEAVAGNPTDVKHIAIKIGPEGSILGCDIA 66
Db 2 KAVIADGQNGVEVISDAPKPTPEKGEFLGVRVAFNPDWKTLYNASEIKGTGGTDFV 61
Qy 67 GTVVKLGPNASTDLKVGDTGTFGVHGSQTDPKNGAFAYARVVPPLFY--KSNLTHSTA 124
Db 62 AVVEDVGEGVDNRKYIGATVSGWAPG--PLDGSNAAWREYITLDVNLVYFVFKNITPS-- 117
Qy 125 DEISEGPKNFESAASLPVSLTTAGVSLCHHLG-----SKMEWHPSPQHTHP 172
Db 118 -----QAATLPLTTTASQGLNQTLGLPLPPTDGSKNSAQQKW----- 155
Qy 173 LLIWGGATAVGQQLIQVAKHINAYTKIVTVASKKHEKLLKSYGADDVFDYHDAGVIRQIK 232
Db 156 VLWWSGSSVQYVVLAAH--AGYKVIATCSPHNFDWIKKLGADFTVDYDHPNVVEIK 213
Qy 233 SKYPNLOHVIDAVGSEDSIPAYKVTADSLPATLLEVP-----MTIESIPPEI 281
Db 214 K-----ATDDSV--FYGDAASFETSTLAVKAFSSKVKDGLNILLSPPSP 259
Qy 282 RKONVKIDITLYRASQEI-LIGATRPASPEYHEATVFKVPINPHANGDIIHNNIK 340
Db 260 RSE-VKLIIGIDYSLFRNEFFNFKTEPQASVDHA-VEYVKLTGWLQEGVIIPNRVK 317
Qy 341 VFSNGLDDVPALTGEGIKGKNKVKYVARL 370
Db 318 EFDGGLQAIKALREFASGKHSVKFVRI 347
RESULT 2
Q8J0F9 PRELIMINARY; PRT; 362 AA.
AC Q8J0F9
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Enoyl reductase.
GN MLCG.
OS Penicillium citrinum.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
OX NCBI_TaxID=5077;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=22162865; PubMed=12172803;
RA Abe Y., Suzuki T., Ono C., Iwamoto K., Hosobuchi M., Yoshikawa H.;
RT "Molecular cloning and characterization of an ML-236B (compactin)
biosynthetic gene cluster in Penicillium citrinum.";
RL Mol. Genet. Genomics 267:636-646(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC MEDLINE=2227327; PubMed=12242508;
RA Abe Y., Suzuki T., Mizuno T., Ono C., Iwamoto K., Hosobuchi M.,
Yoshikawa H.;
RT "Effect of increased dosage of the ML-236B (compactin) biosynthetic
gene cluster on ML-236B production in Penicillium citrinum.";
RL Mol. Genet. Genomics 268:130-137(2002).
DR EMBL; AB072893; BAC20562.1; -.
SQ SEQUENCE 362 AA; 39600 MW; CDB2CC5EFBED0E0F CRC64;
Query Match 15.3%; Score 294; DB 3; Length 362;
Best Local Similarity 26.9%; Pred. No. 2.8e-14;
Matches 104; Conservative 66; Mismatches 160; Indels 56; Gaps 15;
Qy 3 VPTQKXAVIEGDK--AVVKTDVSVPELKEGTALVKVEAVAGNPTDVKHIAIKIGPEGI 60
Db 11 IPPVKQITLTVNDKDEVIWDDAPTCLKPADQYVYRIHAVAVNPSDTK-MRGDFATPFAC 69
Qy 61 LGCDIAGTVVVLGPNASTDLKVGDTGTFGVHGSQTDPKNGAFAYARVVPPLFYKSNLT 120
Db 70 LGTDYAGTVVAVG-SLITHVKVGDRVFGAGNEMCPRTPEQAGAFSQT----- 115
Qy 121 HSTADIESEGPV-----KNFESAASLPVSLTTAGVSLCHHLGSKMEWHPSPQHTHPL 173
Db 116 -----ITRGRIWAKIPDSMTWEAASLPAGISTTGLAM-KLGLMPLPYSETKPSKTYV 168
Qy 174 LIIWGGATAVGQQLIQVAKHINAYTKIVTVASKKHEKLLKSYGADDVFDYHDAGVIRQIKS 233
Db 169 LIYGSSTATATIAMQFMR-LSGYTPIAT-CSHKNFDLAKNGAEVFDYRDADCAQKIRD 226
Qy 234 -KYPNLOHVIDAVGSEDSIPAYKVTADSLPATLLEVPMTIESIPBE-IRKONVKIDIT 291
Db 227 YTRNNLAYALDCIINVESTCYKATGRAGRY-----VALNPFPEHAATKMTSDWT 280
Qy 292 LLYRASQOEILLGATRPASPEYHEATVFKVPINPHL-----NNGDIHNMKVKFS 343
Db 281 L-----GPTIFGEGSTWPA-PYGCEASEE-VRLFGTGLQVASRLVEEDKLYHHPLRVID 333
Qy 344 NGLDDVPALTGEGIKGKNKVKYVAR 369
Db 334 GGLEQVQKQGMETVRNGELSGEKIVVR 359
RESULT 3
Q94110 PRELIMINARY; PRT; 339 AA.
AC Q94110
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Putative quinone oxidoreductase (Fragment).
OS Fragaria ananassa (Strawberry).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Rosales; Rosaceae; Rosoideae; Fragaria.
OX NCBI_TaxID=3747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Chandler;
RA Lopez-Raez J.A., Redondo-Navado J., Medina-Escobar N.,
Caballero-Repullo J.L., Munoz-Blanco J.;
```



```
QY 340 KVFSGNGLDDVPALTEGIGKGNKXKXVRL 370
DB 330 RVQGGFDHKGQMLVRKGLSGEKLVLRL 360

RESULT 6
Q8Y9B9 PRELIMINARY; PRT; 313 AA.
AC Q8Y9B9, 2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DE 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical protein lmo0613.
GN LMO0613.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-e / Serovar 1/2a;
RX MEDLINE=2157279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Ruiniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussauguet O.,
RA Etienne J.-D., Faihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kretz J., Kuhn M., Kunst F., Kurapat G.,
RA Madueno E., Maitounam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL591976; CAC98691.1; --
DR LlistList; LMO0613; --
DR InterPro; IPR002085; Adh_zn_family.
DR Pfam; PF00107; adh_zinc; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 313 AA; 33940 MW; 0F1012C59A23621D CRC64;

Query Match 13.8%; Score 264; DB 16; Length 313;
Best Local Similarity 29.08; Pred. No. 4.4e-12;
Matches 102; Conservative 56; Mismatches 120; Indels 74; Gaps 18;

QY 8 KAVIIE---GDKAVKTDVSVPELKEGTALVKVEAVAGNPTDWK---HTAYKIGPE-GS 59
DB 2 KAVVIENGGKEELKEKEVAMPKAGNQVIVKEAATSINPIDWKLREGYLKQMWDFPI 61

QY 60 ILGCDIAGTVVKGPNASTDLKYGDTGFGFVHGASQTDPKNGAFAEYARYVPLFYKSNL 119
DB 62 ILGWDVAGVISEVGEVY-TDWKVGDEYF-----ARPETTFGTAEYATV-----DDHL 109

QY 120 THSTADISEGPKVKNPESAASLPVSLTTAGVSLCHLGSXMEWHPSPQHTPLLTWGGA 179
DB 110 LAPLEGIS-----FDEAASIPLAGUTAWQALFDH--AKL-----QKGEKVLHAGA 154

QY 180 TAVGQQLIQVAKHINATYKIVTAVSKKHEKLLSYGADDDVFDYHDAGVIEQIKSKYPLNQ 239
DB 155 GGVGTALAIQLAKRAGA--EVITTASAKNHLLSLGADQVIDYKEV---NFKDVLSDID 208

QY 240 HVDAVGSDESIEAYKVTADSLPATLLEVPMTIESIPERKDNVKIDITILLY-RASG 298
DB 209 VVFTMGQGIIE-TDSYDVLKEG-TGRLVISIV-----GISNEDRAKEKNVTATGILQPNG 261

QY 299 QEI-----LLGATRPP-----ASPEYHEATVVFVKFIN 326
DB 262 EQLKELGKLANKTIKPIVGAT-FPPEKGVGFDHALSETHAVGKIVISFN 312

RESULT 7
Q945P3 PRELIMINARY; PRT; 386 AA.
AC Q945P3, 2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DE 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE AT1G23740/F508.27.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shinn P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,
RA Banh J., Bowser L., Carninci P., Dale J.M., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,
RA Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis cDNA clones.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Meyers M.C., Shinn P., Banh J.,
RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis ORF clones.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF411799; AAL06488.1; --
DR EMBL; AY094032; AAM16188.1; --
DR InterPro; IPR002085; Adh_zn_family.
DR Pfam; PF00107; adh_zinc; 1.
SQ SEQUENCE 386 AA; 40986 MW; 4DD19133C9D5D84 CRC64;

Query Match 13.6%; Score 261; DB 10; Length 386;
Best Local Similarity 32.2%; Pred. No. 1e-11;
Matches 88; Conservative 37; Mismatches 104; Indels 44; Gaps 11;

QY 2 SVPTTKQAVIIEG---DXAVKTDVSVPELKEGTALVKVEAVAGNPTDWKHTAYKI--- 54
DB 73 SIPKKNKAWVYSDYGGVDVLLKLESNIVVPEIKEDQVLIKVAAALNFPVDAKRQGFKAT 132

QY 55 -GPEGILGCDIAGTVVKGPNASTDLKYGDTGFGFVHGASQTDPKN-GAFAEYARYVPP 112
DB 133 DGPLTPVPGDVAGVVKVKG-SAVKDLKEGDEVYANVSEKALEGPKQFGLSLAETAAVEK 191

QY 113 L--FYKSNLTHTADEISEGPKVKNPESAASLPVSLTTAGVSLCHLGSXMEWHPSPQHT 170
DB 192 LLALKPKNI-----DFAAAGLPLAETADEGLV-----RTEFSAG----- 227

QY 171 HPLIINGGATVQQLIQVAKHINATYKIVTAVSKKHEKLLSYGADDDVFDYHDAGVIEQ 230
DB 228 KSLVLNAGGAGVSLVQLAKHYGASKVAATASTEKLELVRLSGADLAIDYTKEN-IED 286

QY 231 IKSXPENLQHVDAVGSDES---IPEAYKVTA 259
DB 287 LPDKY---DVVFDAGMCDKAVKVIKEGGKWA 316

RESULT 8
Q92E39 PRELIMINARY; PRT; 313 AA.
AC Q92E39, 2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DE 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
```


KW Complete proteome.
SQ SEQUENCE 311 AA; 34091 MW; 23166D912B5FD248 CRC64;
Query Match 11.4%; Score 219; DB 16; Length 311;
Best Local Similarity 24.9%; Pred. No. 1.2e-08;
Matches 97; Conservative 48; Mismatches 130; Indels 114; Gaps 16;
QY 8 KAVIEG---DKAVVKTDSVPELKEGTALVKVBAVAGNPTDWK-----HIAYKI 54
DB 2 KAVIENYGHADLHEQDPKPTINDNQILTEQVATSNIPIDMKLREGYLKDGFFFPFI 61
QY 55 GPEGSLIGCDIAGTVVKGPNASTDLKVGDFGVHGAASQTDPKNGAFAYARYVPLP 114
DB 62 -----ILGWSAGVVAEVGKRVN--QFVGDRVRSRPATTQA-----GYAEFVAV----- 105
QY 115 YKSNLTHSTADEISEGPKNFESAASLPVSLTTAGVSLCHLGLSKMWHPSPTQHTPL 174
DB 106 -----ESRVAKIPEN--VSFEAAASVPLAGLTAWQCLVDF-----SAIKGDKVL 149
QY 175 IWGATAVGQQLIQVAKHINAYKIVTVVASKKHEKLKSYGADVDVYHDAGVIEQIKSK 234
DB 150 IHAGSGGVGSPAIQFAKHFGAY--VATTASGKNSLVKELGADRFINY-----K 196
QY 235 YPNLQHVIVDAVGSSEDSPEAKVKTADSLPATLLEVPWMTIESIPEETRKONVKIDITILY 294
DB 197 EENFNEVL-----EDYDIVDTLGGDILE-----QSF--EVLKEGKGL----- 232
QY 295 RASQOEILLGATRPASPEYHEATVKVKE-----INPHLNGDIIHWNK 340
DB 233 -----VSTAGNPDESKAKEGKIGAKGLFMLENGKOLSEVADLLENGKRVKSVIGH 281
QY 341 VFSNGLDDVPALTEGKEGKKNKVKYVAR 369
DB 282 TF-----PLTEQGLREAHQLSETHAR 303

RESULT 14

Q8UK00 PRELIMINARY; PRT; 353 AA.
AC Q8UK00;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Zinc-binding oxidoreductase.
GN ATU5324 OR AGR PAT 466.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OG plasmid AT.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Scubel J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Kura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. S.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Rao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
C58.";
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gatung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmlel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,

Flanagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,
Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328(2001).
DR EMBL; AE008954; AAL46012.1; -.
DR EMBL; AE007901; AAK90698.1; -.
DR InterPro; IPR002085; Adh_zn_family.
KW Pfam; PF00107; adh_zinc; 1.
KW Plasmid; Complete proteome.
SQ SEQUENCE 353 AA; 36922 MW; 618FBE7361FFALAS CRC64;
Query Match 11.4%; Score 218; DB 16; Length 353;
Best Local Similarity 29.9%; Pred. No. 1.8e-08;
Matches 80; Conservative 36; Mismatches 106; Indels 46; Gaps 10;
QY 8 KAVIEGDKAVVKTDSVPEL-----KEGTALVKVBAVAGNPTD-----WKHI 50
DB 13 EAMMIEMKAVRULHFEGGPEVLSYEAPRVAASGEVLVRVHAAGINPPDLYLDRGRTL 72
QY 51 AYKIGPEGSS---ILGCDIAGTVVKGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAY 107
DB 73 PPEWQPEPTPLILGTDSGVAAIGDGYSA--ESVGDVEVPAMVRFPEDLMOGSGAYAYV 131
QY 108 RVPPLFYKSNLTHSTADEISEGPVK-NFESAASLPVSLTTAG---VSLCHLGLSKMB-- 161
DB 132 TV-----PASELALKPAGIDHIQAAGAPMSLLTAWQFLVDLGHDAFPNPFQSF 178
QY 162 WHPSPTQHTPLIINGATAVGQQLIQVAKHINAYKIVTVVASKKHEKLKSYGADVD 221
DB 179 RHPSPILQGTVLVNGAGGVGHAIQLAKWGAH--VIAVAGSGKHEALLRALGADQIID 236
QY 222 YHDAGVIEQIKSKVYPNLQHVIVDAVGSSE 249
DB 237 YTKTAA-----ETAEDVDLVIDAVGGSN 260

RESULT 15

O85702 PRELIMINARY; PRT; 313 AA.
ID O85702;
AC O85702 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative oxidoreductase (Putative zinc-binding oxidoreductase).
GN SC00780 OR 3SCF60.12.
OS Streptomyces lividans, and
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycetaceae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1916; 1902;
RN [1]
RP SEQUENCE FROM N.A.
RX SPECIES=S.lividans; STRAIN=1326, AND AJ100;
RA Altenbuchner J.;
RT "Amplifiable element AUD4 from Streptomyces lividans 66.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.coelicolor; STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.coelicolor; STRAIN=A3(2);
RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=S.coelicolor; STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;


```
Db 109 ---HLVDERLVGHAPGSV-DATHAALPLVFLTAWEALFDBL-----PVTQSTGTGLL 157
Qy 175 IWGATAVGQQLIQVAKHINAYTTIVTVASKHEKLLKSYGADVDVHDAGVIEQIKSK 234
Db 158 VLGGAGVPSALIQALRALTNLRVVATASRESTFWVREMGANDVIDHH-----HDLREQ 212
Qy 235 YPNLOHVUDA--VGSEDSI-----PEAYKVTDAD 260
Db 213 ISGVDFVFSWTTGREELTAVMSPQSHLVLD 245

RESULT 23
Q9L2A5
ID Q9L2A5 PRELIMINARY; PRT; 339 AA.
AC Q9L2A5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative zinc-binding oxidoreductase.
GN SC05417 OR SC8F4.21.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RA "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomycetes coelicolor A3(2) chromosome."
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA Rabinowitch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wetzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomycetes
RL coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL939123; CAB70647.1; -.
DR HSSP; P28304; 1QOR.
DR InterPro; IPR002085; Adh zn family.
DR DR00107; adh_zinc; 1.
DR PROSITE; PS01162; QOR_ZETA_CRYSTAL; 1.
KW Complete proteome.
SQ SEQUENCE 339 AA; 34913 MW; ABA43FB244A5C7C8 CRC64;
```

```
Query Match 9.8%; Score 188; DB 16; Length 339;
Best Local Similarity 28.5%; Pred. No. 3.4e-06;
Matches 71; Conservative 30; Mismatches 90; Indels 58; Gaps 10;
```

```
Qy 14 GDKAVVKTDSVPPELKEGTALVKVEAVAGNPTDWKHIAYKI--GPEGSILGCDIAGTVVK 71
```

```
Db 41 GPEVRLTEPRPAPGEGEILVAHAAGLNTDFKRALKIFLPPPLTLGWDVSGTVE 100
Qy 72 LGPNASTDLKVGDTGFG---FVHGASQTDPKNGAFARVYVPPFLPKYKSNLTHSTADEIS 128
Db 101 TGFQV-TLFRPGDEVGMLPYPHG-----HGSABYV----- 131
Qy 129 EGPVKNFES-----AASLPVSLTTAGVSLCHLGSXKWEHPSTPQHTHPLLIWGGA 179
Db 132 TGPAPAPAAKPAAGTGHVQAAALPLAALTAAQALVDTAGLRA-----GQVLIHAAA 182
Qy 180 TAVGQQLIQVAKHINAYTTIVTVASKHEKLLKSYGADVDVHDAGVIEQIKSKYNLQ 239
Db 183 GGVGHVLAQVIAKRGAGH--VTGTASAPKHAFLRELGAACVDRSADFTD--TEERY---D 236
Qy 240 HVIDAVGSE 248
Db 237 VLDTLGGG 245

RESULT 24
Q9KEB8
ID Q9KEB8 PRELIMINARY; PRT; 322 AA.
AC Q9KEB8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Quinone oxidoreductase.
GN BQ0935.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RL halodurans and genomic sequence comparison with Bacillus subtilis."
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001510; BAB04654.1; -.
DR InterPro; IPR02085; Adh zn family.
DR Pfam; PF00107; adh_zinc; 1.
KW Complete proteome.
SQ SEQUENCE 322 AA; 34771 MW; B9D16F204BC5CE2E CRC64;

Query Match 9.8%; Score 187.5; DB 16; Length 322;
Best Local Similarity 27.0%; Pred. No. 3.4e-06;
Matches 79; Conservative 44; Mismatches 83; Indels 87; Gaps 14;

Qy 8 KAVIIE--GKAVVKT-DVSVPELKEGTALVKVEAVAGNPTD-----WKHIAY 52
Db 2 KAVVYEQGDPSEVLKVMVPKPTIGPTDVLINVKASGTPNVDVYFRKGIRQVEQPHIP 61
Qy 53 KIGPEGSILGCDIAGTVVKLGPNASTDLKVGDTGFGVHGASQTDPKNGAFARVYVPP 112
Db 62 -----FDVAGEIVEIGSDIE-NWQVGRVWA-----TNIK-GASAEYALIP 102
Qy 113 LFYKSNLTHSTADEISSEGPVKNFESASLPVSLTTAGVSLCHLGSXKWEHPSTPQHTHP 172
Db 103 LLFP-LPKSV-----SYEEGAALAMTFAHLSL-----FDRGLQGET 141
Qy 173 LLIWGATAVGQQLIQVAKHINAYTTIVTVASKHEKLLKSYGADVDVHDAGVIEQIK 232
Db 142 VLIYGGGAVGHAALQAKRAGA-TVITTAGDREKGEIAKAGADQVIFYKE----- 192
Qy 233 SKYPNLQHVLDVAVGSEDSIPEAYKVTDADSLPATLLEVVWPTIE-SIPEEIRKD 284
Db 193 -----ESVYDAVGK-----ATNGQGVPLIDMSUSENAQD 223
```


SQ SEQUENCE 331 AA; 36851 MW; A3A1F1P4EE177A61 CRC64;
Query Match 9.2%; Score 176; DB 17; Length 331;
Best Local Similarity 25.5%; Pred. No. 2.7e-05;
Matches 80; Conservative 51; Mismatches 107; Indels 76; Gaps 16;
QY 21 TDVSVPELKEGTALVKVAVAGNPTDWHI--AYKIGPEGSILGCDIAGTVVVKLGPNASTD 79
DB 18 TDYKNPEIGNEHVLIRVKLAGVNPVDYVYVERLKNPIPHIPGVFSGEVVKVGDHVK- 76
QY 80 LKVGD--TGFGFV-----HGASQTPDKNGAFAYARVYPPPLPYKNLT 120
DB 77 VSVGDRVTIYGRIFDGTCDMAGVETVCRNGRIGVDANGWAEYVAVBEKYVFKLPNE 136
QY 121 HSTADEISEGPKVFNFESAASLPVSLTTAGVSLCHLGSKMEWHSTPQHTPLIIGGAT 180
DB 137 YT-----WEMGSSLTVAAALTA-----YHALKEAQLSPS-----QTLVVFQASG 174
QY 181 AVGOQLIQVAKHINAYTKIVTVASKGHEKLKSYGADDVFDYHDAGVIEQIK-----SKYP 236
DB 175 NTGMLFVLQKKFGA--KVIAVRK---SWLRGYGADFVDYNE--VVEEKVEITNGKMA 227
QY 237 NLOHVIDAVGSE--DSIPEAYKVTADSLP-ATLLEVVPMTIESIPBEIRKDNKVIDITLL 293
DB 228 DV--VINSLGQFQWDFKPSVLGVGRKLVFTGLT-----GNNVKVDLSQL 270
QY 294 YRASQOEILLGATR 307
DB 271 Y--SKHISILGNVR 282

RESULT 30
Q8UKD5 PRELIMINARY; PRT; 358 AA.
ID Q8UKD5
AC Q8UKD5;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Zinc-binding dehydrogenase.
GN ATU5188 OR AGR_FAT_262.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OG Plasmid AT.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo X., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingley S.V., Tomb J.F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
C58.";
RT Science 294:2317-2323 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
Agrobacterium tumefaciens C58.";
RT Science 294:2323-2328 (2001).

DR EMBL; AE008941; AAL45877.1; ALT_INIT.
DR EMBL; AE007889; AAK90558.1; -
DR InterPro; IPR002085; Adh_zn_family.
DR Pfam; PF00107; adh_zinc; 1.
KW Plasmid; Complete proteome.
SQ SEQUENCE 358 AA; 38909 MW; 929D84D9D9ED11C5 CRC64;
Query Match 9.1%; Score 174.5; DB 16; Length 358;
Best Local Similarity 25.2%; Pred. No. 4e-05;
Matches 98; Conservative 50; Mismatches 148; Indels 93; Gaps 18;
QY 2 SVPTTQ-----KAVIIE---GKAVVKTIDVSPPELKEGTALVKVAVAGNPTDW--KHIA 51
DB 21 SLFSIEQEPKMSKLSKSGGPEAFDLVEVPKPLPKAGQVLVRVHATSIINPLDYQVRRGD 80
QY 52 YK-IGPEGSILGCDIAGTVVVKLGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAYARVY 110
DB 81 YRDLVPLPAITGHDVSGVVEATGFGV-TWPAFGDE---VWYTPQIFDGFSGSYAEYHVN 135
QY 111 PPLFYK--SNLTHSTADEISEGPKVFNFESAASLPVSLTTAGVSLCHLGSKMEWHSTPQ 168
DB 136 ENTIIGRKPSSLTHL-----EASLSLVGGTAMEALVSRALRV----- 173
QY 169 HTHPLLIWGATAVGOQLIQVAKHINAYTKIVTVASKGHEKLKSYGADDVFDYHDAGVI 228
DB 174 -GESILIHGGAGGVHVAIQVAKAIGA--KVYTTVREENFEFAPSRSVGADVVIDYRKEDYV 230
QY 229 EQIKSKYPNL--OHVIDAVGSEDSIPEAYKVTADSLPATLLEVVPMTIESIPBEIRKDNV 286
DB 231 AAIMRETEGLGVDDVDFTLGGE-----TLSHSPKVLQAF--GRVVSIVDIARPN- 278
QY 287 KIDITILYRASGOEILLGATRPASPPEYHEATVKFVKFNPINLNGDIHMMNKKVFSN-- 344
DB 279 -----LIEAWGR-----NASYH-----FVFTRONQKLNELNVLVERGQL 313
QY 345 -----GLDDVPALTEGIGKEGKNVK 365
DB 314 RPHVGAVYSIADLP-LAHALLEKPNNGLR 341

Search completed: January 10, 2004, 10:07:16
Job time : 95 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2004, 05:10:48 ; Search time 4369 Seconds
(without alignments)
10421.702 Million cell updates/sec

Title: US-10-081-644-1
Perfect score: 1113
Sequence: 1 atgcaggttccaccactca.....agtatgtgcaggttataa 1113

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database :

GenEmbl.*

1: gb_be.*

2: gb_htg.*

3: gb_in.*

4: gb_on.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_ey.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1113	100.0	1113	6	AX524811	Sequence
2	1113	100.0	1113	6	BD174421	Novel eno
3	502.2	45.1	509	6	AX524824	Sequence
4	502.2	45.1	509	6	BD174428	Novel eno
5	359	32.3	1134	6	AX524815	Sequence
6	359	32.3	1134	6	BD174423	Novel eno
7	359	32.3	29634	8	YSCLC9122	Saccharomyc
8	357.8	32.1	1145	6	AX524813	Sequence
9	357.8	32.1	1145	6	BD174422	Novel eno
10	357.8	32.1	1631	6	AX536730	Sequence
11	357.8	32.1	1944	8	SCYNL134C	271410 S.cerevisia
12	357.8	32.1	43481	8	SCCHXIV43	246843 S.cerevisia
13	347.8	31.2	316613	8	SCCHRII1	X59720 S.cerevisia
14	345.2	31.0	1122	6	AX524817	Sequence
15	345.2	31.0	1122	6	BD174424	Novel eno
16	252.4	22.7	1083	11	CNS06FEH	AL966303 T7 end of
17	103.6	9.3	942	11	YSCDC91A	AL024117 T7 end of
18	103.2	9.3	5355	8	AX536956	Saccharomyc
19	71	6.4	1086	6	AX536956	Sequence
20	53.2	4.8	1506	8	CCTOXD	X92391 C.carbonum
21	48.6	4.4	212955	2	AC136676	Rattus no
22	48.4	4.3	2000	6	AX655393	Sequence
23	48.2	4.3	109375	8	AC003672	Arabidops
24	47.6	4.3	84550	3	PFMAL1P2_3	Continuation (4 of
25	47.4	4.3	13844	1	AE007561	AE007561 Clostridi
26	47.4	4.3	92357	9	HSJ106901	AL049633 Human DNA
27	47.4	4.3	119082	2	AC008057	Human DNA
28	47.4	4.3	169494	3	AC007630	Homo sapi
29	47.4	4.3	254449	3	AB014817	Homo sapi
30	47	4.2	84118	2	AC096684	Plasmid
31	47	4.2	179553	2	AC024253	Takifugu
32	46.4	4.2	154071	3	AC115598	AC024253 Homo sapi
33	46.2	4.2	5537	3	AP465308	AC115598 Dictyoste
34	46	4.1	110000	2	PFMAL8P1_09	AF466146 Melipona
35	45.8	4.1	14422	3	AP466146	Continuation (10 o
36	45.6	4.1	341050	3	PFAS29357	AF466146 Melipona
37	45.4	4.1	9553	5	AF134321	AF134321 Diaseotic
38	45.4	4.1	303414	1	AE015938	AF134321 Diaseotic
39	44.8	4.0	2922	5	LFL308116	AE015938 Clostridi
40	44.8	4.0	164487	2	AL954708	AE015938 Clostridi
41	44.8	4.0	242302	2	AC131200	AJ308116 Lampetra
42	44.6	4.0	11327	1	AE006787	AL954708 Danio rer
43	44.6	4.0	163243	9	AL513304	AC131200 Rattus no
44	44.2	4.0	74154	2	AC096997	AE006787 Sulfolobu
45	44.2	4.0	186935	2	AC022322	AL513304 Human DNA
46	44.2	4.0	236263	2	AC098600	AC096997 Takifugu
47	44	4.0	244845	2	AC126846	AC022322 Homo sapi
48	43.8	3.9	157695	9	AL136442	AC098600 Rattus no
49	43.8	3.9	213544	2	BX510939	AL136442 Rattus no
50	43.8	3.9	221756	2	AC094893	AL136442 Rattus no
51	43.8	3.9	232974	2	AC114193	BX510939 Danio rer
52	43.8	3.9	253001	3	AB014834	AC094893 Rattus no
53	43.8	3.9	299650	1	AP004603	AC114193 Rattus no
54	43.8	3.9	325350	1	AP004823	AP004603 Oceanobac
55	43.6	3.9	242516	2	AC106523	AP004823 Staphyloc
56	43.6	3.9	243203	2	AC090437	AC106523 Rattus no
57	43.6	3.9	250541	1	AE017004	AC090437 Mus muscu
58	43.6	3.9	294769	2	AC098907	AE017004 Bacillus
59	43.4	3.9	110000	2	PFMAL7P1_05	AC098907 Rattus no
60	43.4	3.9	94434	5	AL929250	Continuation (6 of
61	43.2	3.9	110000	2	PFMAL7P1_08	AL929250 Zebrafish
62	43.2	3.9	184090	10	AC122304	Continuation (9 of
63	43.2	3.9	191613	10	AC122304	AC122304 Mus muscu
64	43.2	3.9	243490	2	AC131536	AC122304 Mus muscu
65	43.2	3.9	249995	3	AE014840	AC131536 Rattus no
						AE014840 Plasmid

66	43.2	3.9	313096	2	AC131879	AC131879 Rattus no
67	42.8	3.8	2000	6	AX655393	AX655393 Sequence
68	42.8	3.8	116696	3	PFMAL3P3	Z98547 Plasmodium
69	42.6	3.8	4111	10	MMU278580	AJ278580 Mus muscu
70	42.6	3.8	125026	10	AC091616	AC091616 Rattus no
71	42.6	3.8	178273	2	AC005308	AC005308 Plasmodiu
72	42.6	3.8	183420	2	AC116417	AC116417 Mus muscu
73	42.6	3.8	185814	2	AC116888	AC116888 Mus muscu
74	42.6	3.8	198550	2	AC140782	AC140782 Mus muscu
75	42.6	3.8	219955	2	AC127414	AC127414 Mus muscu
76	42.6	3.8	241327	2	AC127126	AC127126 Rattus no
77	42.6	3.8	251075	2	AC132995	AC132995 Rattus no
78	42.6	3.8	269050	1	AP000987	AP000987 Sulfolobu
79	42.6	3.8	349751	3	PFMAL4P3	AL035476 Plasmodiu
80	42.4	3.8	171730	2	AL929558	AL929558 Danio rer
81	42.4	3.8	291490	1	AE017030	AE017030 Bacillus
82	42.2	3.8	206836	2	AC129047	AC129047 Rattus no
83	42.2	3.8	222603	2	AC125755	AC125755 Rattus no
84	42.2	3.8	234097	2	AC120493	AC120493 Rattus no
85	42.2	3.8	248793	2	AC103080	AC103080 Rattus no
86	42.2	3.8	301050	1	AP003130	AP003130 Staphyloc
87	42.2	3.8	333321	3	AC116986	AC116986 Dictyoste
88	42.2	3.8	343590	1	AP003359	AP003359 Staphyloc
89	42	3.8	96573	2	AC136672	AC136672 Rattus no
90	42	3.8	110000	2	AC111624_2	Continuation (3 of

RESULT 1	AX524811	AX524811	1113 bp	DNA	linear	PAT 21-NOV-2002
LOCUS	AX524811	Sequence 1 from Patent EP1236796.				
DEFINITION	AX524811					
ACCESSION	AX524811					
VERSION	AX524811.1	GI:25169907				
KEYWORDS						
SOURCE	Kluyveromyces lactis					
ORGANISM	Kluyveromyces lactis					
REFERENCE	Eukaryota; Fungi; Ascomycota; Saccharomycetaceae; Kluyveromyces.					
AUTHORS	Yamamoto,H. and Kimoto,N.					
TITLE	Novel enzyme reductions isolated from kluyveromyces lactis, methods for producing same, and methods for selectively reducing a carbon-carbon double bond of an alpha, beta-unsaturated ketone using the reductions					
JOURNAL	Patent: EP 1236796-A 1 04-SEP-2002;					
FEATURES	Daicel Chemical Industries, Ltd. (JP)					
SOURCE	Location/Qualifiers					
	1..1113					
	/organism="Kluyveromyces lactis"					
	/mol_type="genomic DNA"					
	/db_xref="taxon:28985"					
CDS	1..1113					
	/note="unnamed protein product"					
	/codon_start=1					
	/protein_id="CAD57033.1"					
	/db_xref="GI:25169908"					
	/translation="MSVPTQKAVIIIEGDKAVKTVDSVPPELKEGTALVKVEAVAGNP					
	TDWKHAIKIGPEGSILGCDIAGTVKVPNGASTDLKVGDTGFGFVHGASQDTPKNGA					
	FAEYRVVPLFVKYKSLNSTADEISEGPKVNFESASLPVSLTTAGVSI-CHHLSKMK					
	EWHPSTQHPHLLINGGATVAGQQLIQVAKHINAYKIVTVASKHEKLLKSYGADD					
	VFDHDAVIEQKSKPTPNLQHVDAVGSSEDSIPEAYKVTADSLPATLLEVPMTIES					
	IPEIRKDNKYKIDITLLYRASGOEILLGATRFPSPEYHEATYKFKVFNPHNLNGDI					
	HHNKKVPSNGLDVPALTEIGIKGKKNKYVARL"					
BASE COUNT	343 a	210 c	234 g	326 t		
ORIGIN						

Query Match	100.0%;	Score 1113;	DB 6;	Length 1113;
Best Local Similarity	100.0%;	Pred. No. 6e-238;		
Matches 1113;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;


```

Db      1081 AACAAAAATGTTAGTATGTTGCCAGGTATAA 1113
|||||
RESULT 2
BD174421
LOCUS      1113 bp      DNA      linear      PAT 18-MAR-2003
DEFINITION Novel enone reductase, process for producing the same, and method
            for selectively reducing carbon-carbon double bonds of alpha,
            beta-unsaturated ketone using the same.
ACCESSION BD174421
VERSION    BD174421.1 GI:29120109
KEYWORDS   JP 2002247987-A/1.
SOURCE     Kluyveromyces lactis
ORGANISM   Eukaryota; Fungi; Ascomycota; Saccharomycetaceae; Saccharomycetes;
            1 (bases 1 to 1113)
REFERENCE  Yamamoto,H. and Kimoto,K.
AUTHORS    Novel enone reductase, process for producing the same, and method
TITLE      for selectively reducing carbon-carbon double bonds of alpha,
            beta-unsaturated ketone using the same
JOURNAL    Patent: JP 2002247987-A 1 03-SEP-2002;
            DAICEL CHEMICAL INDUSTRIES LTD
COMMENT    OS Kluyveromyces lactis
            PN JP 2002247987-A/1
            PD 03-SEP-2002
            PF 23-FEB-2001 JP 2001049363
            PI HIROAKI YAMAMOTO,KUNIHIRO KIMOTO
            PC C12N15/09,C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12N9/02,C12P7/
            PC 26//
            CC Novel enone reductase, process for producing the same, and CC
            method for
            CC selectively reducing carbon-carbon double
            bonds of alpha,beta-
            CC ketone using the same
            FH Key Location/Qualifiers
            FT CDS Location/Qualifiers
FEATURES   source
            1..1113
            /organism="Kluyveromyces lactis"
            /mol_type="genomic DNA"
            /db_xref="taxon:28985"
BASE COUNT 343 a 210 c 234 g 326 t
ORIGIN
Query Match 100.0%; Score 1113; DB 6; Length 1113;
Best Local Similarity 100.0%; Pred. No. 6e-238;
Matches 1113; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      1 ATGTCAGTTCMAACCACTCAAAAAGCCGTATCATTTGAAGGTGACAAAGCTGTTGTTAAA 60
Db      1 ATGTCAGTTCMAACCACTCAAAAAGCCGTATCATTTGAAGGTGACAAAGCTGTTGTTAAA 60
Qy      61 ACAGATGTCCTCAGTTTCAGAAATTAAGGAGGTACAGCTTTGGTGAAGGTTTGAGGCTGTT 120
Db      61 ACAGATGTCCTCAGTTTCAGAAATTAAGGAGGTACAGCTTTGGTGAAGGTTTGAGGCTGTT 120
Qy      121 GCTGGTAACCACTGATTGGAAGCATATTGCTTTATAAGATTGTCAGAAAGGTTTCAATT 180
Db      121 GCTGGTAACCACTGATTGGAAGCATATTGCTTTATAAGATTGTCAGAAAGGTTTCAATT 180
Qy      181 CTAGATGTGACATTCGTGGTACAGTTGTCAAACTTGGACCAAAATGCTAGTACTGACTTG 240
Db      181 CTAGATGTGACATTCGTGGTACAGTTGTCAAACTTGGACCAAAATGCTAGTACTGACTTG 240
Qy      241 AAGGTTGAGATACCGGTTTCGGTTTGGTTTGTTCACCGTCTTCCCAACAGATCCCTAAAAT 300
Db      241 AAGGTTGAGATACCGGTTTCGGTTTGGTTTGTTCACCGTCTTCCCAACAGATCCCTAAAAT 300
Qy      301 GGTGCATTTGCTGAATATGCCAGGTTTATCCACCTTTGTTTACAGAGTAACTTAACT 360
|||||
301 GGTGCATTTGCTGAATATGCCAGGTTTATCCACCTTTGTTTACAGAGTAACTTAACT 360
|||||
361 CACTCAACTGCTGATGAAAATTTCTGAAGGCCCTGTGAAGAACTTCGAATCTGCTGCATCA 420
361 CACTCAACTGCTGATGAAAATTTCTGAAGGCCCTGTGAAGAACTTCGAATCTGCTGCATCA 420
421 TTGCCAGTTTTCGTTGACAACTGCTGCTGTTAGTTTGTGTCATCAGTTCGAGCTCAAAAATG 480
421 TTGCCAGTTTTCGTTGACAACTGCTGCTGTTAGTTTGTGTCATCAGTTCGAGCTCAAAAATG 480
481 GAATGGCAGCCATCTTACCCCGCAACATACATCATCATCATTTATTGTTGGGGTGGTGTACA 540
481 GAATGGCAGCCATCTTACCCCGCAACATACATCATCATCATTTATTGTTGGGGTGGTGTACA 540
541 GCAGTGGGTCAACAACTTAATCCAAAGTTGCCAAACATATCAATGCTTATCTAAGATTGTA 600
541 GCAGTGGGTCAACAACTTAATCCAAAGTTGCCAAACATATCAATGCTTATCTAAGATTGTA 600
601 ACTGTTGCTTTCTAAAAAGCATGAAAAGCTTTTAAAGTCTTTATGCTGCTGATGATGCTTTT 660
601 ACTGTTGCTTTCTAAAAAGCATGAAAAGCTTTTAAAGTCTTTATGCTGCTGATGATGCTTTT 660
661 GACTATCATGATGACAGCGTTATTGAGCAGATCAAATCGAAGTATCCAAACCTTGCAACAT 720
661 GACTATCATGATGACAGCGTTATTGAGCAGATCAAATCGAAGTATCCAAACCTTGCAACAT 720
721 GTTATTGACGCTGTGGGAAGCATAGTATCCCGAGGCCCTATAAAGTCACAGCAGAT 780
721 GTTATTGACGCTGTGGGAAGCATAGTATCCCGAGGCCCTATAAAGTCACAGCAGAT 780
781 AGTCTACCTGCCACATTTATTAGAAGTGGTTCCAAATGACCAATGAAAAGCATTCCTGGAAGAA 840
781 AGTCTACCTGCCACATTTATTAGAAGTGGTTCCAAATGACCAATGAAAAGCATTCCTGGAAGAA 840
841 ATCAGAAAAGATAATGTTAAAAATTGATATTAATTTGTTGTTGATCGTGATCTGTCAGAA 900
841 ATCAGAAAAGATAATGTTAAAAATTGATATTAATTTGTTGTTGATCGTGATCTGTCAGAA 900
901 ATTCTATTGGGTGCAACAGATTTTCCTGCTAGTCCAGAAATATCATGAAGCCACAGTTAAA 960
901 ATTCTATTGGGTGCAACAGATTTTCCTGCTAGTCCAGAAATATCATGAAGCCACAGTTAAA 960
961 TTGTTTAAAGTTTATAAATCCACACCTTTAACACGGTGATATCCATCATATGAATATTTAAA 1020
961 TTGTTTAAAGTTTATAAATCCACACCTTTAACACGGTGATATCCATCATATGAATATTTAAA 1020
1021 GTTTTCAGCAACCGCTTTAGATGATGTCCTGCTAGTCCAGCTCTCACTGAAGGTATATAAAGAGTTAA 1080
1021 GTTTTCAGCAACCGCTTTAGATGATGTCCTGCTAGTCCAGCTCTCACTGAAGGTATATAAAGAGTTAA 1080
1081 AACAAAAATGTTAGTATGTTGCCAGGTATAA 1113
1081 AACAAAAATGTTAGTATGTTGCCAGGTATAA 1113
|||||
RESULT 3
AX524824
LOCUS      509 bp      DNA      linear      PAT 21-NOV-2002
DEFINITION Sequence 14 from Patent EP1236796.
ACCESSION AX524824
VERSION    AX524824.1 GI:25169918
KEYWORDS   Kluyveromyces lactis
SOURCE     Kluyveromyces lactis
ORGANISM   Eukaryota; Fungi; Ascomycota; Saccharomycetaceae; Saccharomycetes;
            1
REFERENCE  Yamamoto,H. and Kimoto,N.
AUTHORS    Novel enone reductases isolated from kluyveromyces lactis, methods
TITLE      for producing same, and methods for selectively reducing a
            carbon-carbon double bond of an alpha, beta-unsaturated ketone
            using the reductases

```


Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.

1
Yamamoto, H. and Kimoto, N.
Novel enone reductases isolated from *Kluyveromyces lactis*, methods
for producing same, and methods for selectively reducing a
carbon-carbon double bond of an alpha, beta-unsaturated ketone
using the reductases
Patent: JP 1236796-A 5 04-SEP-2002;
Daicel Chemical Industries, Ltd. (JP)

JOURNAL
Location/Qualifiers
1. .1134
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/db_xref="taxon:4932"
1. .1134
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAD57035.1"
/db_xref="GI:25169912"
/translation="MAQVAIPETMKAVVIEDKAVKGIPIPELBERGFVLIKTLAVA
GNPTDWAHIDYKIGPQSIILGCDAGOI VKLGPVNPDPISIGDYIYGFTHGSSVRP
SNGAPAEYSALSTVAYKSPNELKPLGEDVLPAGVRSLSGVATIPVSLTAGLVITY
NLGDLWESETPGRKPIILMGATVQSLIOLANKUNGFTKIIIVASRKHKLK
EYGADELFDHIDVBOIKHKYNNISYLVDCVANQDTLOQVYKCAADKQDATIVELK
NLTEBNYKKNRQNVTDIIRLYSIGGHEVPFGNITLPADSEARAAKAIKFIKFINPK
INDQIRHIPVRVYKGLCDVPHILKDIKYGNSGEKLVAVLN"

BASE COUNT 351 a 225 c 245 g 313 t
ORIGIN

Query Match 32.3%; Score 359; DB 6; Length 1134;
Best Local Similarity 58.9%; Pred. No. 1.1e-69;
Matches 658; Conservative 0; Mismatches 450; Indels 9; Gaps 2;

Qy 5 CAGTTCACACCTCAAAAGCCGTCATCTGAAGGTGACAAAGCTGTGTTAAACAG 64
Db 14 CAATTCAGAAACCATGAAGCTGTCTCATTTGAAGACGGTATGTTTAAAGAGG 73
Qy 65 ATGTCCTAGTTCACAGAAATTAAGAGGGGTACAGCTTGGTGAAGCTTGAGCTGTGCTG 124
Db 74 GCATTCCTCATCTGAAATTTGAAGAGGATTCGTATTGATTAACACACTGCTGTGCTG 133
Qy 125 GTAACCAACTGATTTGAAGCATATGCTTTAAGATTTGTCACAGAGGTTCAATTTCTAG 184
Db 134 GTAACCCCACTGATTTGGGCACATTTGACTACAAGATCGGGCCCTCAAGGATCTATTCTGG 193
Qy 185 GATGTGACATTTGCTGTGATCAGTTGTCAAACTTGACCAATGCTAGTACT---GACTTGA 241
Db 194 GATGTGATGCTGTGGCCAAATTTGTCAAATTTGGGCCCCAGCTGTCAATCTTAAAGACTTTT 253
Qy 242 AGGTTGGAGATACCGGTTTTCGTTTGTTCACGGTCTTCCCAACAGATCTTAAAGATG 301
Db 254 CTATCGGTGATATATTATGGGTTTCATTCACGATCTTCCGTAAGTTTCCTTCCCATG 313
Qy 302 GTGCATTTGCTGAATATGCCAGGTTTATCCACCTTTGTTTTTACAA-----GAGTAACT 355
Db 314 GTGCTTTTGTGTAATATTCTGCTATTCAACTGCTGGTTGCTCACAATCAACCAATGAC 373
Qy 356 TAACTCACTCACTGCTGATGAATATTTCTGAAGCCCTGTGAAGAACTTCGAATCTGCTG 415
Db 374 TCAAAATTTTGGGTGAGGATTTCTTACCTGCGGCCCTGTGAGGTCTTTGGAAGGTGTAG 433
Qy 416 CATCATTTGCCAGTTTCGTTGACAACTGCTGTGTTAGTTTGTGTCATCACTTGGGCTCAA 475
Db 434 CCATATCCCACTGCTGCTGACCAAGCCGCTGTGTTGACCTATAACTTTGGGCTTGG 493
Qy 476 AAATGGATGGCAACCCATCTACCCCGCAACATACTCATCTCAATTTATGTTGGGGTGGT 535
Db 494 ACCTGAGTGGGACCATCAACCCCAACAAAGAAAGGCCCATCTTTATTATGGGGCGGTG 553
Qy 536 CTACAGCAGTGGGTCAACAACTAATCCAAAGTTGGCAACATATCAATGCTTTATCTAAGA 595
Db 554 CAATGTCAGTAGGTGAGTGGCTCATCCAAATTTAGCCAAATTAATGAATGGCTTCAACAGA 613

596 TTCTAACTGTGTCTTCTTAAAGCATGAAAAGCTTTTAAAGCTTTATGCTGCTGATG 655
614 TCATTGTGTGGCTTCTCGGACGACGAAAACCTTTTGAAGAAATATGCTGCTGATGAAT 673
656 TCTTGACTATCATGATGACGGGTATTATGAGCAGATCAAAATCGAAGTATCCAAACCTGC 715
674 TATTTGATTATCATGATATTGACGTGTGTAGAACAAAATTTAAACAACAAGTACAAATATCT 733
716 AACATGTTATTGACGCTGTGGGAAGCAAGATAGTATCCCGAGGCTTATAAAGTCACAG 775
734 CGTATTATGTCGACTGTGTCGGAATCAAGATACGCTTCAACAGGTGTACAAATGTGCGG 793
776 CAGATAGTCTACCTGCCACATTTATTAGAAGTGGTTCCAAATGACCAATTTGAAAGCATTCCTG 835
794 CCGATAAACAGGATGCTACAAATTTGTTGAATTTAAAAAATTTGACAGAGAAACGTCACAAA 853
836 AAGAAATTCAGAAAAGATAATGTTAAATTTGATATTACTTTTGTGTCGTCATCTGGTC 895
854 AAGAGAACAGGACACAAAACGTTACTATTGACATAAAGGGCTATATTCAATAGGTGGCC 913
896 AAGAAATTCATTTGGGTGCAACAGATTTCTCTAGTCCAGAAATATCATGAAAGCCACAG 955
914 ATGAAGTACCAATTTGGAAACATTTACTTTACCAAGCCGACTCAGAAAGCTAGGAAAGCTGCAA 973
956 TTAATTTGCTTAAGTTTATAAATCCACACCTTAACCAAGGTGATATCCCATCATATGAATA 1015
974 TAAATTTATCAATTTCAATTTCAAGATTAATGATGCAAAATTCGCCATATTCCAG 1033
1016 TTAAGTTTTCAGCAACGGCTTAGATGATGTCCAGCTCTCACTGAAGGTATTAAGAAG 1075
1034 TAAGGGTCTATAAGAACGGGCTTTGTGATGTCTCATATCTCTAAAGACATCAAAATATG 1093
1076 GTAAGAACAAAATTTTAAGTATGTTGCCAGGTTATA 1112
1094 GTAAGAACTGTGTGAAAAAACTCGTTGCGGTATTAA 1130

RESULT 6
BD174423
LOCUS
DEFINITION
BD174423 1134 bp DNA linear PAT 18-MAR-2003
Novel enone reductase, process for producing the same, and method
for selectively reducing carbon-carbon double bonds of alpha,
beta-unsaturated ketone using the same.
BD174423
BD174423.1 GI:29120111
VERSION
JP 2002247987-A/3.
KEYWORDS
Saccharomyces cerevisiae (baker's yeast)
Saccharomyces cerevisiae
ORGANISM
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE
1 (bases 1 to 1134)
Yamamoto, H. and Kimoto, N.
Novel enone reductase, process for producing the same, and method
for selectively reducing carbon-carbon double bonds of alpha,
beta-unsaturated ketone using the same
Patent: JP 2002247987-A 3 03-SEP-2002;
DAICEL CHEMICAL INDUSTRIES LTD
COMMENT
OS Saccharomyces cerevisiae (yeast)
PN JP 2002247987-A/3
PD 03-SEP-2002
PF 23-FEB-2001 JP 2001049363
PI HIRAKI YAMAMOTO, KUNIHICO KIMOTO
PC C12N15/09, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12N9/02, C12P7/
26//
PC C12N15/09, C12R1:645, C12N15/00, C12N5/00, (C12N15/00, C12R1:645)
CC Novel enone reductase, process for producing the same, and CC
method for
CC selectively reducing carbon-carbon double
bonds of alpha, beta-
unsaturated
CC ketone using the same
FH Key Location/Qualifiers

```

FEATURES             FT  CDS           Location/Qualifiers
     source            1. 1134
                        /organism="Saccharomyces cerevisiae"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:4932"
BASE COUNT           351 a 225 c 245 g 313 t
ORIGIN
Query Match          32.3%; Score 359; DB 6; Length 1134;
Best Local Similarity 58.9%; Pred. No. 1.1e-69;
Matches 658; Conservative 0; Mismatches 450; Indels 9; Gaps 2;

Qy 5 CAGTTCACCAACCACTCAAAAGCCGTCATCATTTGAAGGTGACAAAGCTGTGTTAAACAG 64
    |||||
Db 14 CAATTCAGAACCACTGAGGCGTGTGCTATGAGACGGTGAAGCCGGTGTAAAGAGG 73

Qy 65 ATGTCCTCAGTTCAGAAATTAAGGAGGTACAGCCTTTGGTGAAGTTGAGCGTGTGCTG 124
    |||||
Db 74 GCAATTCCTCCTGAAATTTGAAGAGGATTCGTATTGATTAAAGACACTGCTGTGCTG 133

Qy 125 GTACCCCACTGATGGAAGCATATTGCTTATAGATTGTCGAGAGGTTCAATTCCTAG 184
    |||||
Db 134 GTAACCCCACTGATGGAAGCATATTGCTTATAGATTGTCGAGAGGTTCAATTCCTAG 193

Qy 185 GATGTGACATTTGCTGTACAGTGTGCAAACTTGACCAAAATGCTAGTACT--GACTTGA 241
    |||||
Db 194 GATGTGATGCTGTGCGCAAAATTTGCAAAATTTGCGCCCGCTGTGCTCAATCCTAAAGACATTT 253

Qy 242 AGTTGGAGATPACCGGTTTCGGTTTGTTCACGGTGTCTCCAAACAGATCCTAAAAATG 301
    |||||
Db 254 CTATCGGTGATTATATTATGCGTTCATTCACGGATCTTCGGTAAGTTTCTTCCCAATG 313

Qy 302 GTGCAATTTGCTGATATGCGAGGTTTATCACCTTTGTTTTTACAA-----GAGTAAC 355
    |||||
Db 314 GTGCTTTTGTGGAATATCTGCTATTTCAACTGTGTTGCTTACAAATCAACCAATGAAC 373

Qy 356 TAACCTCACTCAACTGCTGATGAAATTTCTGAAGCCCTGTGAAGAACTTCGAATCTCCTG 415
    |||||
Db 374 TCAATTTTGGGTGAGGATGTTCTACTCGCGGCCCTGTGAGTCTTTGAGAGGTGTAG 433

Qy 416 CATCAATGCCAGTTTCGTGTGACAACTGCTGTGTGTAGTTGTGTGTCATCATTGGGCTCAA 475
    |||||
Db 434 CCACATATCCCACTGCTCACTGACCAACAGCCGGCTTGGTGTGACCTATAAATCTGGGCTGG 493

Qy 476 AAATGGAATGCCACCCATCTACCCGCAACATATCTCATCTCAATTTATGATTTGGGGTGG 535
    |||||
Db 494 ACCTGAAATGGAGGCCATCAACCCCAAGAAAGGCCCACTCTTATTATGGGGCGGTG 553

Qy 536 CTACAGCAGTGGGTCAACAACTAATCAAGTTGCCAAACATATCAATGCTTATATAAGA 595
    |||||
Db 554 CAATCGAGTAGGTGAGTGGTCAATCAGTCAATAGGCAATTAATGAAATGGCTTCAACAGA 613

Qy 596 TTGTPAACCTGTGCTTCAAAAGATGAAAGCTTTTAAAGTCTTATGTCGTGATGATG 655
    |||||
Db 614 TCATTGTGTGGCTTCTCGGAAGCACGAAACCTTTTGAAGAATATGGTGTGATGAAT 673

Qy 656 TCATTGACTATCATGATGACGGCTTATGAGCGAGATCAATCGAAGTATCCAAACCTGC 715
    |||||
Db 674 TATTGATTAATCATGATATGACGTGTAGAACAAATTAACACAAAGTACAACAATATCT 733

Qy 716 AACATGTTATTGACGCTGTGGGAAGCAAGATAGTATCCCGAGGCCCTATAAAGTCACAG 775
    |||||
Db 734 CGTATTATTAGTCAGTGTGCGAATCAGATAGCGTTCAACAGTGTACAATGTGCGG 793

Qy 776 CAGATGCTACCTGCCACATTAATGAAAGTGGTTCCAAATGACCAATGAAAGCATTCCTG 835
    |||||
Db 794 CCGATAAACAGGATGCTACAAATTTGAAATTTTGAAGAAATTTGACAGAGAAACGTCACAAA 853

Qy 836 AAGAAATCAGAAAGATAATGTTAAATTTGATTAATTTGTTGTTGATCGTCACTGGTC 895
    |||||
Db 854 AAGAGAACAGGAGACAAACGTTACTATTGACATAATTAAGGCTATATTCAATAGGTGGCC 913
```

```

RESULT 7
YSL9122/c
LOCUS           29634 bp  DNA  linear  PLN  01-AUG-1997
DEFINITION     Saccharomyces cerevisiae chromosome XII cosmid 9122.
ACCESSION     U22383 Y1338
VERSION       U22383.1 GI:2264349
KEYWORDS
SOURCE        Saccharomyces cerevisiae (baker's yeast)
ORGANISM
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE
1 (bases 1 to 29634)
  Johnston,M., Hillier,L., Riles,L., Albermann,K., Andre,B.,
  Ansoerge,W., Benes,V., Bruckner,M., Delius,H., Dubois,E.,
  Dusterhoft,A., Entian,K.D., Floeth,M., Goffeau,A., Hebling,U.,
  Heumann,K., Heuss-Neitzel,D., Hilbert,H., Hilger,F., Kleine,K.,
  Kotter,P., Louis,E.J., Messenguy,F., Mewes,H.W., Miosga,I.,
  Mostl,D., Muller-Auer,S., Nentwich,U., Obermaier,B., Piravandi,E.,
  Pohl,T.M., Portetelle,D., Purnelle,B., Rehmman,S., Rieger,M.,
  Rinke,M., Rose,M., Scharfe,M., Scherens,B., Scholler,P.,
  Schwager,C., Schwarz,S., Underwood,A.P., Urrestarazu,L.A.,
  Vandenbol,M., Verhasselt,P., Vierendeels,P., Voest,M., Volckaert,G.,
  Voss,H., Wambutt,R., Wedler,E., Wedler,H., Zimmermann,F.K.,
  Zollner,A., Hani,J. and Hoheisel,J.D.
  The nucleotide sequence of Saccharomyces cerevisiae chromosome XII
  Nature 387 (6632 Suppl), 87-90 (1997)
9169871
PUBMED
2 (bases 1 to 29634)
  Kirsten,J.
  The sequence of S. cerevisiae cosmid 9122
  Unpublished (1994)
REFERENCE
3 (bases 1 to 29634)
  Waterston,R.
  Direct Submission
  Submitted (08-MAR-1995)
  4 (bases 1 to 29634)
  Cherry,J.M.
  Direct Submission
  Submitted (18-JUL-1997) Saccharomyces Genome Database
  5 (bases 1 to 29634)
  Cherry,J.M.
  Direct Submission
  Submitted (28-JUL-1997) Saccharomyces Genome Database
  On Jul 18, 1997 this sequence version replaced gi:717070.
Sequenced by:
  Genome Sequencing Center
  Department of Genetics, Washington University,
  St. Louis, MO 63110, USA
Curated by:
  Saccharomyces Genome Database
  URL: http://genome-www.stanford.edu/
  e-mail: yeast-curator@genome.stanford.edu
Neighboring Sequence:
  This sequence includes nucleotides 1-22200 of cosmid 9122. This is
```

the rightmost cosmid of chromosome 12. The cosmid on the left is 9324. This sequence overlaps the right telomeric sequence, accession number 273327, starting at nucleotide 28299.

FEATURES	Location/Qualifiers		
	1..29634	/organism="Saccharomyces cerevisiae"	
		/mol_type="genomic DNA"	
		/strain="S288C (AB972)"	
		/db_xref="taxon:4932"	
		/map="XII"	
	51..7937	/gene="YLR454W"	
	51..7937	/gene="YLR454W"	
		/codon_start=1	
		/evidence=not experimental	
gene		/product="Ylr454wp"	
		/protein_id="AAB64718.1"	
		/db_xref="GI:717075"	
		/translation="MMFPINLVLYKMLFAVTFWLSCKILLRKLGINITWINLFKLE ICGLSLEDGTVRLKSVRAVERKLFIKGLRIDSKSSTNDLHKELPREERTFIETP BNGSGFISKILSOLYWLNGVTIILEDOLVNDITIERKGFPLSIDNSKH1KSLRP DSPLKLNWGQIIADAI FVNTNLLIGIMNPKOGLQVGLDGLDNLINPNNLLN LFTKENVDMSKELQLRADTTKANEEUDEDIAKDDLVYAMEKFDVRIKPLKE MNTVDKLOIKDPLTNHPELGNKYISYNVLVSINFNENTNFRNEMPQYTLIFER DSEPKFSINARINYLNLNRKQSHAKQLKIIENPNVSI FGSNTLFSQKFRLSNNLH AKELNAPNIGNISLTIDMPNVSFIKFLSNLKVFTSSCPNKKILKENSXKVF LTRRVLFDYKCFPLPINMKFTLDOPKFVINDKDLIIGKFSVFMISHSKRYTLGN NLSAEKETOHIIFYESHWNVLLDMKQLHIIKQKYEHTILRVDSTAIERKQVLLPDI LCSANADITLDLSELPMVMSLWNLDSOLANVEENYKFEYKFPASLQNMK AECNSMAKRLKELPSPDFMQQLPDDFYIKINIRIDISHTIGARSVFMPPRVFSSV DSQSKDLIDGKRLKNTVEKQIALFGDKTQWNGKIGSNHATMVRSGOLTWFSDN KONPNHKSITADDDISTSDATEVNLWNLNLVNDITTSIIIGETPEVSELSKTVS KVNSLSKLPDTESSFNESDSKILQINHSGTSVMSLSIFLAUSGHTLNQIFG HCHQNRQSKTKQIFALASESSKSKIKWQLKLEINFSEISIIQIALPNG LRTKPESTITVTKNCTISVSGQYFMVMSFTQPNFWMRCINGFKYKWIHIDL KQMKLNLQNWEKLPKLSAITLNDWSHFSIPIHFEMFKIIDSIPITFKSIKQMLYSL KLOEDNKRISDILSKQNAKEMKDDFEADNAILKRTGLWAKDGKRLKRSATDS EIPLTAALNKGKDDRPDRTQFISPEIENAYNTLLANESDSWIKRKYKVERRE FDKNSFLWGFIDYTKLPIDINKKVLFPSTPFLMNLIIENIDIDIRPSCGIENIPN FIHDVGKGVNPTIYSIIMPHLDAPKFEVRLRDYPLFVSIPLSSTSQSKETIPM RIYGDGFIETMDLQSDRELRTLFVPLIPSVTVENTDRYYSLFPRTMTSAKIFDNLN EIMSNHTRTVWGSYQPAIQOTWQCLDNFSKPLDPSVKLGFWDKTRYLFGHKINIV WKRKGFSEILGAKSPYMLGSGSAGFIIVGFGNVLKCNEDNPKKFLSCSADKVFH SINPYFAKPLVWSPSTNMFIPNODDTNMQRYASFYLLNTTSSQKADKEIMGK SPIKYGIKLSGWTLMGILFERLGSNLNERTTESKGYLTRCLNPIYVODLUSKSDS YAGFRSDFIHMSFGLSNSNSAYNAMQLSPNAGDAQFVWPKFSGNPPVRRGFLGLQ SISPKFGHELYTISYHADVSPLFINMYHNDADQILRNKYLEVAFAGLKAYSKSHFI MDLQRKEVLTEYQAGLNVRVRVWMLKFLAGDVVQDVIDRTVSGFSLKLYTEEKD AeyDIIPDNMSLIDITDQDAFFINPDNYPKIKIMPFAPSPQFAYQKRSYDKYQV DPKTCPIPTPDNRVSHGCTLGHVSLRTDLVEKRVTVLAKPREKQEGIRKNSAGV SEENLDDLKSNLSVENABELLKPOKIFQKHAEGQTEQPFHDSNLNLKRLTKLK QFBEHFFIFNLWKNEDASAIKFFFYANLSENFASLASGKLEFEDVIRQREMT DDTTSMEAIPEGDKANTTKQCHSCDDTEFTENLNIIEKNITQLSCDINKIHHKF FVQFITQIOLTSLNEACVLSVSPFMKLTLEFDANTTNTYMODIFLRKHILFG NANAFLPKKDYQRELYFGSSSYGODKKEQWPMWGLGLEPSPALKKAVNRIS ALLHQKLAIPSAKDSLKDIEDNICGYQVQNVQNSDEYMLTKMALKFLIYVEP EDELKYLIEKLIIGYDIYDQTRKFVNDLHDSQILAVVEKLELFRSLDDIDGKL DLNHNHNRHQHLLRYLIRKVTNSGNNINRTLWNIKVNETILHLDKIDKPLD IYAVKLNFORQIHLRKLNTVYQMMQIFDLGNNVNYHCILGPIITSSGNDTVGLAS DVLQVQITPDVDPVGGIKVQKVVETLSLTIKLEBDRLNKLFEWLSKLVDGNG DDDGASSIPDMVSSSECKIIEFSEDISDFNEMLKRSDYMDVEDLKLNSKELCI YKKGQRNLANTNVNFPFTRLNSQTLRTVDTLLALKVKLIKVLIRKHTGRFIGNKL KRNSKNKIADDTSPKLQTLTYNSPEELR"	
		complement (8124..8197)	
		/gene="t1(AAU) L2"	
	tRNA		complement (8124..8197)
			/gene="t1(AAU) L2"
	repeat_unit		/product="tRNA-Ile"
			/note="codon recognized: AUU"
		complement (8357..8685)	
		/note="delta element YLRdelta27"	

Query Match 32.3%; Score 359; DB 8; Length 29634;
Best Local Similarity 58.9%; Pred. No. 9.1e-70;
Matches 658; Conservative 0; Mismatches 450; Indels 9; Gaps 2;

Qy 5 CAGTTCACACCACTCAAAAGCCGCTCATATGTAAGGTGACAAAGCTGTTGTTAAACAG 64
Db 16930 CAATTCCAGAACCACTGAAGGCTGCTGTCATTGAAGACGGTAAAGCGGTTGTTAAAGAG 16871

Qy 65 ATGCTCAGTTCACAGATTAAAGGAGGTACAGCCTTGGTCAAGTTGAGGCTGTTGCTG 124
Db 16870 GCATTCCCATCTCGAATTGGAAGAGGATTGCTGATTGATTGAACACACTCGCTGTTGCTG 16811

Qy 125 GTAAACCAACAGTGTGAAGCATATGCTTATAGATTGGTCCAGAAAGTTTCAATTTCTAG 184
Db 16810 GTAAACCAACAGTGTGGCACACATTGACTACAGATCGGGCTCAAGGATCTATTTCTG 16751

Qy 185 GATGTACATGCTGGTACAGTTGTCAAACTGTCACCAATGTCAGTAAGTACT---GACTGA 241
Db 16750 GATGTGATGCTGCTGGCCAAATTTGCAATTTGGGCCAGCTGTCAATCTCTAAAGACTTTT 16691

Qy 242 AGTTGGAGATACCGTTTCGGTTTCTTCACTGCTTCCCAACAGATCCTTAAAGATG 301
Db 16690 CTATCGGTGATTAATTATGCGGTTCATTACAGGATCTTCGTAAGGTTTCTCTTCCAAAG 16631

Qy 302 GTGCATTGCTGTAATGTCAGAGGTTTATCCACCTTTGTTTACAA-----GAGTAACT 355
Db 16630 GTGCTTTTGTGATATATCTGCTATTTCACTGCTGGTGTGCTACAAATCACCACATGAAC 16571

Qy 356 TAACCTACTCACTGCTGATGAATTTCTGAAGCCCTGTGAAGAACTTCGAATCTGCTG 415
Db 16570 TCAATTTTGGGTGAGGATGTTCTACCTGCGGCCCTGTGAGTCTTTGGAAGGTGTAG 16511

Qy 416 CATCAATTGCCAGTTTCTGACACTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 475
Db 16510 CCACTATCCCAGTGTCACTGACCAACGCGGCTTGGTGTGACCTATAACTTGGGCTGG 16451

Qy 476 AATGGAATGGCACCCTATCCCGCAACATACTCATCTCATTTATGTTTGGGTTGGTG 535
Db 16450 ACCTGAAGTGGAGCCATCAACCCCAAGAAAGGCCCTCTTATTAATGGGCGGTG 16391

Qy 536 CTACAGCAGTGGGTCACAACTAATCAAGTGTGCCAAACATCAATGCTTATTAATAAGA 595
Db 16390 CAATGTCAGTAGTTCAGTCTGCTATCCAAATAGCAATAAATGAAATGAAATGCTTCAACA 16331

Qy 596 TTGTAACCTGTGCTCTAAAGAGCATGAAAGCTTTTAAAGTCTTATGGTCTGATGATG 655
Db 16330 TCATTGTTGGCTTCTCGGAAGCAGCAAAAACCTTTTGAAGAATATGGTCTGATGAAT 16271

Qy 656 TCTTTGACTATCATGATGCAAGGCTTATGAGCAGATCAATCGAAGTATCCAAACCTGC 715
Db 16270 TATTTGATTATCATGATTAATGAGTGTAGACAAATTAACACAGTACACATATCT 16211

Qy 716 AACATGTTATGAGCTGTGGGAAGCAAGATAGTATCCCGAGGCTTATAAGTACACAG 775
Db 16210 CGTATTTAGTCGACTGTGCGAATCAAGATACGCTTCAACAAGTGTACAAATGTGCG 16151

Qy 776 CAGATAGTCTACCTGCCACATTAATGAGTGGTTCATGACCATTTGAAGCATTCCTG 835
Db 16150 CCGATAACAGGATGCTACAAATTTGTAATTTAAATAATTTGACAGAAAGAAACGTCAAA 16091

Qy 836 AAGAAATCAGAAAGATAATGTTTAAATTCATATTAATTTGTTGTTATCGGTGATCTGGTC 895
Db 16090 AAGAAACAGAGACAAACGTTACTATTAAGATTAATTAAGGCTATTAATAGGTGGCC 16031

Qy 896 AAGAAATCTTATTTGGGTGCAACAGATTTCTCTGCTAGTCCAGAAATATCATGAAGCCACAG 955
Db 16030 ATGAAGTACCAATTTGGAACATTAATTTACCGCGACTCAGAAGCTTAGGAAAGCTGCA 15971

Qy 956 TTAATTCGTTAGTTTATTAATTCACACCTTTAAACAGGTGATATCCATCATATGAATA 1015
Db 15970 TAAATTTTATCAATTCATCAATCCAAAGATTAATGATGGAACAAATTCGCCATATTTCCAG 15911

Qy 1016 TTAAAGTTTTTCAGCAACGGCTTAGATGATGCCAGCTCTCCAGGTGATATAAAGAG 1075
Db 15910 TAAGGCTCTATRAAGAACGGCTTTGATGTTCTCATATCTCTAAAGACATCAATATG 15851

Qy 1076 GTAAACCAAAATGTTTAAGTATGTTGCCAGGTTATA 1112
Db 15850 GTAAGAACTCTGCTGTAAGAACTCGTTGCGGTATTA 15814

RESULT 8
AX524813
LOCUS AX524813 1145 bp DNA linear PAT 21-NOV-2002
DEFINITION Sequence 3 from Patent EP1236796.
ACCESSION AX524813
VERSION AX524813.1 GI:25169909
KEYWORDS Saccharomyces cerevisiae (baker's yeast)
SOURCE Saccharomyces cerevisiae
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE 1
AUTHORS Yamamoto, H. and Kimoto, N.
TITLE Novel enzyme reductases isolated from kluyveromyces lactis, methods for producing same, and methods for selectively reducing a carbon-carbon double bond of an alpha, beta-unsaturated ketone using the reductases
JOURNAL Patent: EP 1236796-A 3 04-SEP-2002;
Daicel Chemical Industries, Ltd. (JP)

FEATURES
source
1. .1145
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/db_xref="taxon:4932"
6. .1136
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD57034.1"
/db_xref="GI:25169910"
/translation="MSASIPETWKAUVIENGKAVUKODIPIPELSEGFVLIKTVAVAG
NPTDWHIDFKIPQGLGCDAGQIVKIGENVDAARPAIDGYIYGVHIGASVRPPS
NGAFAYSAISSETAYKPAEPRLCGKDLPEGLVGLGSLVPSLUTAGMILTHS
FGLDWTFKPSKAQRDPIPLFWGATAVGOMLIQAKLNGFSKIIVASRRHEKLKE
YGADLFDYHDADVI EQIKYNNI PYLVDCVSNFTETQQVYKCAADLDATVVLTV
LTKEDIKEEDRONYSIEGTLLYLGNDVPPFTLPADPEYKEAAIKFKIPNPKI
NDGEIHLPVKYVKNGLDIPOLLDDIKHRSNGEKLVAVLK"

BASE COUNT 343 a 245 c 256 g 301 t

ORIGIN
Query Match 32.1%; Score 357.8; DB 6; Length 1145;
Best Local Similarity 58.8%; Pred. No. 2e-69;
Matches 658; Conservative 0; Mismatches 452; Indels 9; Gaps 2;

Qy 4 TCAGTTCCAAACCACTCAAAAGCCGCTCATATTGAAGGTGACAAAGCTGTTGTTAAACA 63
Db 15 TCGATTCCAGAAACCACTGAAGCCGTTGTCATTGAAATGGCAAGCTGTAGTCAAAACAG 74

Qy 64 GATGCTCAGTTCAGATTAAGGNGGTACAGCCTTGGTGAAGTTGAGCTGTGCT 123
Db 75 GACATTCCTCAATTCCTGAATTTAGAAGAGGATTTGTTCTAATTAAGACTGTCGCGTTGCC 134

Qy 124 GGTAAACCAACCTGATTGGAAGCATATTGCTTATAAGATTGGTCCAGAAAGTTTCAATTCTA 183
Db 135 GGTAAACCTTACGATTGGAAACATATTGATTCAAGATTGGTCCCTCAAGTGCCTCTTA 194

Qy 184 GGATGTGACATTGCTGGTACAGTTGTCAACTTGGACCAAAATGCTAGTACTG---ACTTG 240
Db 195 GCGTGTGATGACGCGGCCCAATCGTAAAGTTGGGCCCAAAATGTTGATGCTGCAGCTTT 254

Qy 241 AAGGTTGGAGATACCGGTTTCGGTTTGTTCACGCTGCTTCCCAACAGATCCTTAAAT 300
Db 255 GCCATTGGGTGATTACATTTATGGGGTTATTACGCGTGTCTTCAAGTGGTCCCTCAAC 314

Qy 301 GGTGCATTTGCTGAATATGCCAGGCTTTATCCACCTTTGTTTATTACAA-----GAGTAC 354

||||| 315 GGTGCTTGGCTGAGTACTGCTGCACTTTCATCCGAGACTGCTTATAAACCAGCAGAG 374
QY 355 TTAACCTCACTCAACTGCTGATGAAATTTCTGAAGGCCCTGTGAAGAACTTCGATCTGCT 414
Db 375 TTTAGATTGTCGGTAAGACAAGCTACCAAGAGGCCCTGTAATCTTTAGAGGGGCA 434
QY 415 GCATCAATGCCAGTTTGGTTGCAACTGCTGGTTAGTTTGTGTCATCACTTTGGGCTCA 474
Db 435 GTATCCCTCCAGTCTCATTGACACAGCGCTGGTATGATCCTTACACATAGTTTGGCTTG 494
QY 475 AAAATGGAATGGACCCATCTACCCGCAACATCTCATCTCATTTATGATTGGGGTGGT 534
Db 495 GACATGACATGGAAGCCCTCAAAGCGCAAGAGATCAACCCATCTTATTTGGGGTGGT 554
QY 535 GCTACAGCAGTGGTCAACAACCTAACTCAAGTTGGCAACATATCAATGCTTATATAAG 594
Db 555 GCACCTGCTGGCCAGATGCTTATTCATTTGGCAAAAACATAAAGGTTTCAGCAAG 614
QY 595 ATTGTAACCTGTTCTTAAAAAGCATGAAAGCTTTTAAAGTCTTTATGGTGTGATGAT 654
Db 615 ATCATGCTGCTGCTTCTGTAACATGAAATAATTTGTTGAAGAGTACGGTGCAGATGA 674
QY 655 GTCTTTGACATCATGATGACAGCGTTATTTAGCAGATCAAAATCGAAGTATCCAACTG 714
Db 675 CTTTTTGACTACCAAGTCTGACGTTATCGAACAGATAAAAGAAAGTACAAACATTT 734
QY 715 CAACATGTTTATTCAGCTGTTGGCAAGCAAGATAGTATCCCGAGGCTATAAAGTCACA 774
Db 735 CTTACTTGTGACTGTGTTCTCAACACAGAAACTATTCAAGGTGTCAAAATGTGCC 794
QY 775 GCAGATAGTCTACCTGCCACATATTATAGAAAGTGGTTCATGACCATTTGAAGCATTCCT 834
Db 795 GCTGATGACTTTAGACGCTACGGTCTTCAATTGACCGTTTAAACGAAAGATATCAAG 854
QY 835 GAAGAAATCAGAAAGATAATGTTAAATGATATTACTTTGTTGATCTGTCATCTGTT 894
Db 855 GAGGAAGACAGAGGCAAAACGTCAGTATTGAAAGAAACCCCTTCTATATTGATAGGAGT 914
QY 895 CAAGAAATTTCTATTGGTGCACAAAGATTTCTGCTAGTCCAGAAATATCATGAAGCCCA 954
Db 915 AACGAGCTCCATTTGCGACGTTTACTTTGCGACAGACCTCGAATACAAAGGAAGCGCC 974
QY 955 GTTAAATTCGTTAAATTTATAATCCACACCTTAAACAAAGGATATCCATCATATGAAT 1014
Db 975 ATAAATTTTAAATTATCATCATCCAAATCAATGATGCTGTAATCCACCATCCCA 1034
QY 1015 ATTAAAGTTTTCAGCAAGCGCTTAGATGATGTCCTCAGCTCTCACTGAAGGTATAAAGAA 1074
Db 1035 GTGAAAGTTTACAAAGACCGGTTAGATGATATCCACAGTTACTTGTATGATATTAAGCAC 1094
QY 1075 GGTAAACCAACAAATGTTAAGTATGTTGCCAGGTTATAA 1113
Db 1095 GGGAGGAATTTCTGGCAAAAGTTGGTTGCGCTCTTGAAA 1133

RESULT 9
BD174422
LOCUS
DEFINITION
1145 bp DNA linear PAT 18-MAR-2003
Novel enone reductase, process for producing the same, and method
for selectively reducing carbon-carbon double bonds of alpha,
beta-unsaturated ketone using the same.
BD174422
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
BD174422.1 GI:29120110
JP 2002247987-A/2.
Saccharomyces cerevisiae (baker's yeast)
Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
1 (bases 1 to 1145)
Yamamoto, H. and Kimoto, K.
Novel enone reductase, process for producing the same, and method
for selectively reducing carbon-carbon double bonds of alpha,

beta-unsaturated ketone using the same
Patent: JP 2002247987-A 2 03-SBP-2002;
DAICEL CHEMICAL INDUSTRIES LTD
OS Saccharomyces cerevisiae (yeast)
PN JP 2002247987-A/2
PD 03-SBP-2002
PF 23-FEB-2001 JP 2001049363
PI HIROAKI YAMAMOTO, KUNIHITO KIMOTO
PC C12N15/09, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12N9/02, C12P7/
26//
PC (C12N15/09, C12R1:645), C12N15/00, C12N5/00, (C12N15/00, C12R1:645)
CC Novel enone reductase, process for producing the same, and CC
method for
CC selectively reducing carbon-carbon double
bonds of alpha, beta-
CC ketone using the same
CC unesterated
FH Key Location/Qualifiers
FT CDS (6)..(1136).
FEATURES
source
1..1145
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/db_xrefs="taxon:4932"
BASE COUNT 343 a 245 c 256 g 301 t
ORIGIN
Query Match 32.1%; Score 357.8; DB 6; Length 1145;
Best Local Similarity 58.8%; Pred. No. 2e-69;
Matches 658; Conservative 0; Mismatches 452; Indels 9; Gaps 2;
QY 4 TCAGTTCCAACTCAAAAGCCGTCATCATTGAAGGTGACAAAGCTGTGTAAACA 63
Db 15 TCAGTTCCAGAAACCATGAAGCGTTGTCATTGAAATGGCAAGCTGTAGTCAACAG 74
QY 64 GATGCTCAGTTCCAGAAATTAAGAGGGGTACACCTTGGTGAAGGTGAGGCTGTGCT 123
Db 75 GACATTTCAATTCCTGAATTTAGAAAGGATTTGTTCTAATTAAGACTGTGCGCGTTGCC 134
QY 124 GGTAAACCACTGATTGGAGCATATTTGCTTATAAGATTGGTCCAGAGGTTCAATCTA 183
Db 135 GGTAAACCTTACCGATTGGAAACATATTGATTTCAGATTGGTCTCAAGGTGCCCTTTA 194
QY 184 GGAATGACATTCGCTGATGAAATTTCTGAAGGCCCTGTGAAGAACTTCGAATCTGCT 240
Db 195 GGTGATGATGACGCGCGCCAAATCGTAAAGTTGGGCCAAATGTTGATGCTGACGCTTT 254
QY 241 AAGTTGGAGATACCGGTTTCGGTTTTCAGGTTGCTTCCCAAACAGATCTCAAAAT 300
Db 255 GCATTTGGTATTACATTTTATGGGTTATTACGGTGTCTTCAAGTGGTTCCTCAAC 314
QY 301 GGTGCAATTTGCTGAATATGCCAGGTTTATCCACCTTTGTTTACAA-----GAGTAAC 354
Db 315 GGTGCTTTGCTGATGATCTCTGCCATTTTCATCCGAGACTGCTTATAAACCAGCAGAGAG 374
QY 355 TTAACCTCACTCAACTGCTGATGAAATTTCTGAAGGCCCTGTGAAGAACTTCGAATCTGCT 414
Db 375 TTTAGATTGTCGGTAAAGACAAGCTACCAAGAGGCCCTGTAATCTTTAGAGGGGCA 434
QY 415 GCATCAATGCCAGTTTGGTTGACAACTGCTGGTTAGTTTGTGTCATCACTTTGGGCTCA 474
Db 435 GTATCCCTCCAGTCTCATTGACACCGGCTGGTATGATCCTTACACATAGTTTGGCTTG 494
QY 475 AAAATGGAATGGACCCATCTACCCGCAACATCTCATCTCATTTATGATTGGGGTGGT 534
Db 495 GACATGACATGGAAGCCCTCAAAGCGCAAGAGATCAACCCATCTTATTTGGGGTGGT 554
QY 535 GCTACAGCAGTGGTCAACAACCTAACTCAAGTTGGCAACATATCAATGCTTATATAAG 594
Db 555 GCACCTGCTGGTGGCCAGATGCTTATTCATTTGGCAAAAACATAAAGGTTTCAGCAAG 614
QY 595 ATTGTAACCTGTTCTTAAAAAGCATGAAAGCTTTTAAAGTCTTTATGGTGTGATGAT 654

Db 615 ATCATCGTCGTCCTCTCGTAAACATGAAAAATTTGTTAAAGAGTACGGTGCAGATGAA 674
QY 655 GTCTTTGACTATCATGATGAGGGCTTATTGAGCAGATCAAAATCGAAGTATCCAAACCTG 714
Db 675 CTTTTTGACTACACGATGCTGAGCTTATCGAACAGATAAAAGAAAGTACAACACATT 734
QY 715 CAACATGTTATTGACGCTGTGGGAAGCGAAGATAGTATCCCGAGGCTTATAAAGTCACA 774
Db 735 CTTTACTTGGTGGACTGTCTCCAACACAGAACTATTCAACAGGTGTACAAATGTGCC 794
QY 775 GCAGATAGTCTACCTGCCACATTATTAGAGTGTTCCTCAATGACCAATTTGAAGCATTCCT 834
Db 795 GCTGATGACTTAGACGCTACGGTCTGTTCAATTGACCGCTTTTAAACCGAAAAAGATATCAAG 854
QY 835 GAAGAAATTCATGAAAGATATGTTAAATTTGATATTTGTTGTTATCGTGCATCTGGT 894
Db 855 GAGGAGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 914
QY 895 CAAGAAATTCATTTGGGTGCAACAGATTTCTGCTAGTCCAGAAATATCATGAAAGCCACA 954
Db 915 AACGAGCTCCCATTTGGCAGCTTTACTTTGCGCAGCAGACCTGATACAGGAGGAGCGCC 974
QY 955 GTTAAATTCGTTAAGTTTAAATTCACACCTTAACAAACGCGTATCCATCATATGTAAT 1014
Db 975 ATAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 1034
QY 1015 ATAAAGTTTTCAGCAACGCTGTAGATGATGTCAGCTCTCACTGAAGGTATAAAGAA 1074
Db 1035 GTGAAGTTTACAGAAACGGGTGTAGATGATGATGATGATGATGATGATGATGATGATGAT 1094
QY 1075 GGTAAAAACAAAAATTTAAGTATGTTGCCAGGTATTAA 1113
Db 1095 GCGAGGAATTCGCGGAAGTTGTTGCGCTCTTGAA 1133

RESULT 10
AX536730
LOCUS AX536730 1631 bp DNA linear PAT 22-NOV-2002
DEFINITION Sequence 331 from Patent WO02064766.
ACCESSION AX536730
VERSION AX536730.1 GI:25263166
KEYWORDS Saccharomyces cerevisiae (baker's yeast)
SOURCE Saccharomyces cerevisiae
ORGANISM Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1
AUTHORS Contreras, R.H., Eberhardt, I., Luyten, W.H. and Reekmans, R.J.
TITLE Bax-responsive genes for drug target identification in yeast and
fungi
JOURNAL Patent: WO 02064766-A 331 22-AUG-2002;
JANSSEN PHARMACEUTICA N.V. (BE)
FEATURES
Location/Qualifiers
1. 1631
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/db_xref="taxon:4932"
BASE COUNT 477 a 357 c 346 g 451 t
ORIGIN

Query Match 32.1%; Score 357.8; DB 6; Length 1631;
Best Local Similarity 58.8%; Pred. No. 2e-69;
Matches 659; Conservative 0; Mismatches 452; Indels 9; Gaps 2;
QY 4 TCAGTCCCAACCACTCAAAAGCGGTGTCATGTTGAAGGTGACAAAGCTGTTGTTAAACA 63
Db 510 TCAGTCCCAACCACTCAAAAGCGGTGTCATGTTGAAGGTGACAAAGCTGTTGTTAAACA 569
QY 64 GATGCTCTCAGTTCAGAAATTAAGGAGGTGACAGCTTGTGTAAGGTGAGGCTGTGCT 123
Db 570 GACATTCCAATTCCTGAATTTAGAAAGAGGATTTGTTCTAATTAAGACTGTCGCGGTGCC 629
QY 124 GGTAAACCAACTGATTGGAAGCATATTTGCTTATAAGATTGGTCCAGAAAGTTCAATTCTA 183

Db 630 GGTAAACCTCAACGATTGGAAACATATTGATTTCAGAGTTGGTCTCAAGGTGCGCTCTTA 689
QY 184 GATGTGACATTTGCTGGTACAGTTGTCAAACTTGGACCAAAATGCTAGTACTG---ACTTG 240
Db 690 GGTGTGATGTCAGCGCGCCAAATCGTAAAGTTGGGCCAAATTTGTTGATGCTGCACGCTTT 749
QY 241 AAGTTGGAGATACCGGTTTCGGTTTTCAGGTTCTTCCAAACAGATCCTTAAAAAT 300
Db 750 GCCATTGGTGATTAACATTTATGGGTTATTACGGTGTCTTCAAGTGTCTTCAAGTGTCCCTCAAC 809
QY 301 GGTGCATTTGCTCAATATATCCAGGGTTTATCCACCTTTGTTTACAA-----GAGTAAC 354
Db 810 GGTGCTTTGCTGAGTACTCTGCCATTTTCATCCGAGACTGCTTATAAACCAGCCAGAGAG 869
QY 355 TTAACCTCACTCACTGCTGATGAAATTTCTGAAGGCCCTGTGAAGNACTTCAATCTGCT 414
Db 870 TTTAGATTGTGGGTAAAGACAGCTACCAAGGCCCGCTAAATCTTTAGAAAGGGCA 929
QY 415 GCATCATTCGCCAGTTTCTGTTGACAACTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 474
Db 930 GTATCCCTCCAGTCTCATTTGACCGCTGTGATGATCCTTACACATAGTTTGGCTTG 989
QY 475 AAAATGGAATGGCACCCCATCTACCCCGCAACATACTCATCTCATTTATTTGGGTGGT 534
Db 990 GACATGACATGGAAGCCCTCCAAAGCGCAAGAGATCAACCCCATCTTATTTTGGGTGGT 1049
QY 535 GCTACAGAGTGGGTCAACAACTAATCCAAAGTTGCCAAACATATCAATGCTTATCTAAG 594
Db 1050 GCCACTGCTGTTGGCCAGATGCTTATTTCAATTTGGCAAAAAAACTAAACGGTTTCAGCAAG 1109
QY 595 ATTGTAACCTGTTGCTTCTAAAAAGCATGAAAGCTTTTAAAGTCTTTATGCTGCTGATGAT 654
Db 1110 ATCATGCTGTTGCTTCTCTGTAACATGAAAAATTTGTTGAAGAGTACCGTGCAGATGAA 1169
QY 655 GTCTTTGACTATCATGATGACGGCGTTATTGAGCAGATCAAAATCGAAGTATCCAAACCTG 714
Db 1170 CTTTTTGACTACCGATGCTGACGTTATCGAAAGATATAAAGAAAGTACAAACATTT 1229
QY 715 CAACATGTTATTGACGCTGTGGGAAGCGAAGATAGTATCCCGAGGCTTATAAAGTCACA 774
Db 1230 CCTTACTTGGTGGACTGTGCTCTCAACACAGAAACTATTCAACAGGTGTACAAATGTGCC 1289
QY 775 GCAGATAGTCTACCTGCCACATTTATTAGAAGTGGTTTCCAATGACCATTCGAAAGCATTCCT 834
Db 1290 GCTGATGACTTAGACGCTACGGTCTCAATTTGACGGTTTTTAAACCGAAAGATATCAAG 1349
QY 835 GAAGAAATCAGAAAAAGATAATGTTAAATTTGATATTTACTTTTGTGTATCGTGCATCTGGT 894
Db 1350 GAGGAAGACAGGAGGCAAAACGTCAGATTGAAGGAACCTTCTATATTTGATAGGAGT 1409
QY 895 CAAGAAATTCATTTGGGTCAACAGATTTCTGCTAGTCCAGAAATATCATGAAAGCCACA 954
Db 1410 AACGACGTCCTCAATTTGGCAGCTTTACTTTTGGCAGCAGACCTCGAATACAAAGAAAGCGCC 1469
QY 955 GTTAAATTCGTTAAGTTTATAAATCCACACTTAAACAGGTGATATCCATCATATGAAT 1014
Db 1470 ATAAATTTATTAGTTTCAATCAATCAAAATCAATGATGTTGTTGAATCCCAATCCCA 1529
QY 1015 ATTAAGTTTTCAGCAACGGCTTAGATGATGTCAGCTCTCACTCAAGGTATATAAAGAA 1074
Db 1530 GTGAAAGTTTCAAGAAAGGGTTAGATGATATCCACAGCTTACTTGTATGATATTAAGCAC 1589
QY 1075 GGTAAAAACAAAAATTTAAGTATGTTGCCAGGTATTAA 1113
Db 1590 GGGAGGAATTCGTGGCAAAAGTTGTTGCGCTCTTGA 1628

RESULT 11
SCYNL134C/c
LOCUS SCYNL134C 1944 bp DNA linear PLN 11-AUG-1997
DEFINITION S.cerevisiae chromosome XIV reading frame ORF YNL134C.
ACCESSION Z71410 Y13139


```
VERSION 271410.1 GI:1302075
KEYWORDS Saccharomyces cerevisiae (baker's yeast)
SOURCE Saccharomyces cerevisiae
ORGANISM Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 1944)
AUTHORS Mallet, L., Bussereau, F. and Jacquet, M.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1944)
AUTHORS MIPS.
TITLE Direct Submission
JOURNAL Submitted (29-APR-1996) Data collected by MIPS on behalf of the
European yeast chromosome XIV sequencing project. MIPS at the
Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
Martinsried, FRG; E-mail: Mewes@mips.emblnet.org
FEATURES
    Location/Qualifiers
        1..1944
            /organism="Saccharomyces cerevisiae"
            /mol_type="genomic DNA"
            /db_xref="taxon:4932"
            /chromosome="XIV"
            complement(225..1355)
            /note="ORF YNL134c"
            /codon_start=1
            /protein_id="CAA96016.1"
            /db_xref="GI:1302076"
            /db_xref="SWISS-PROT:P53912"
            /translation="MSASIPETKAVIENGKAVVQKIDIPLEBEGFVLKTVAVAG
            NPTDKHIDFKIGQALGCDAGQIVKLGPNVDAARFAIGDYIYVHIGASVRPFS
            NGAPAEYSASSETAYPAREFLRCGLPEGVKSLGAVSLPVLTAGMILTHS
            FGLDMTWKPSKQRDQPIFLWGGATVAGQMLIQLAKLNGFSKIIIVASRHEKLLKE
            YGADLPDYHDADVEIOIKKKYNNIPYLVDVSNTEIIQOVYKCAADLDATVVQLTV
            LTKEDIKEEDRQNVSTIEGLLYLIGNDVPFRGFTLPADPEYKEAIIKFIKINPKI
            NDGEIHIPVKYNGKLDLIPQLDDDIKHGNSGEKLVAVLK"
BASE COUNT 542 a 397 c 420 g 585 t
ORIGIN
Query Match 32.1%; Score 357.8; DB 8; Length 1944;
Best Local Similarity 58.8%; Pred. No. 2e-69;
Matches 658; Conservative 0; Mismatches 452; Indels 9; Gaps 2;
Qy 4 TCAGTTCCAAACCTCAAAAGCGCTCATCATCTGAAGGTGCAAAAGCTGTGTGTAAACA 63
Db 1346 TCGATTCCAGAAACCATGAAAGCGGTGTCTATTGAAATGCGAAGCTGTAGTCAACAG 1287
Qy 64 GATGCTCAGTTCAGAAATTAAGCAGGTACACCTTGGTGAAGGTGAGCGTGTGCT 123
Db 1286 GACATTCCAATTCCTGAATTAGAGAGAGATTTGTTCTTAATTAAGACTGTGCGCGTTGCC 1227
Qy 124 GGTAACCACTGATTGGAAGCATATTGCTTATAAGATTGTCAGAAAGGTTCATTA 183
Db 1226 GGTAACTCTACGATTGGAACATATTGATTTCAAGATTGCTCTCAAGTGCCCTCTTA 1167
Qy 184 GGATGTGACATTCGTGTGACAGTTGTCAAACTTGGACCAATGCTAGTACTG---ACTTG 240
Db 1166 GGCTGTGATGAGCGCGCCAAATCGTAAAGTTGGCCCAATGTTGATGCTGCACGCTTT 1107
Qy 241 AAGTTGGAGATACCGGTTTCGGTTTGTGTCACGGTGTCCCAACAGATCTCAAAAT 300
Db 1106 GCCATTTGGTATTAATTTATGGGGTTATTACGGTGTCTTCAGTGAGGTTCCTCAAC 1047
Qy 301 GGTCATTTGCTGAATATGCCAGGTTTATCCACCTTTGTTTACAA-----GAGTAAC 354
Db 1046 GTGCGCTTTGCTGAGTACTCTGCCATTTTCACGAGACTGCTTATAACACGACAGAG 987
Qy 355 TTAATCACTCACTGCTGATGAAATTTCTGAAGCGCTTGTGAAGAACTTCGAATCTGCT 414
Db 986 TTTAGATTGTCGGTAAAGACAAGCTACCAAGAGCGCCCGTAAATCTTTAGAGGGGCA 927
Qy 415 GCATCATTCGAGTTTCGTTGACAACTGCTGGTGTAGTTTGTGTGATCACTTGGGCTCA 474
Db 926 GTATCCCTCCAGTCTCATTTGACCACGCGTGTATGATCTCTTACACATAGTTTGGCTTG 867
```

```
475 AAATGGAATGGCACCCTCTACCCGCAACATACCTCATCTATTTGATTTGGGGTGGT 534
866 GACATGACATGGAAGCCCTCCAAAGCGCAAGAGATCAACCCATCTTATTTGGGGTGGT 807
535 GCTACAGCAGTGGGTCAACACATTAATCCCAAGTTGCCAACAATATCAATGCTTATTAAG 594
806 GCCACTGCTGTGGCCAGATGCTTATTCATTTGGCAAAAAAATAACCGGTTTCAGCAAG 747
595 ATTGTAACCTGCTCTTAAAAAAGCATGAAAAGCTTTTAAAGCTTTATGCTGCTGATGAT 654
746 ATCATGCTGCTGCTCTCTGTAACAATGAAAATTTGTTGAAAGAGTACGGTGCAGATGAA 687
655 GTCTTTGACTATCATGATGACGCGTTTATTGACGAGATCAAAATCGAAGTATCCAAACCTG 714
686 CTTTTTGGACTACCAAGTCTGACGTTTATCGAAGATATAAAGAAAGTACAAACATTT 627
715 CAACATGTTATTGACGCTGTGGGAAGAGATAGTATCCCGAGGCTTATAAGTCACA 774
626 CTTTACTTGGTGAAGTGTGCTCCAAACACAGAAACTATTCAACAGGTTGTACAAATGTGCC 567
775 GCAGATAGTCTACCTGCCATTTATTAGAGTGTGTTCCCAATGACCATTTGAAGCATTTCT 834
566 GCTGATGACTTAGACGCTACGGTCTCAATTCACCGTTTAAACCGAAAAGATATCAAG 507
835 GAAGAAATCAGAAAAGATATGTTAAATTTGATATTACTTTGTTGATTCGTCATCTGGT 894
506 GAGGAAGACAGGAGGCAAAACGTCAGTATTGAAGGAACCTTCTATATTGATAGGAGT 447
895 CAAGAAATTTCTATTGGGTGCAACAAGATTTCTGCTAGTCCAGAAATATCATGAAGCCACA 954
446 AACGACGCTCCATTTGCGCAGTTTACTTTGCCACGACAGCCTGAATACAAAGGAAGCGGCC 387
955 GTTAAATTCGTTAAGTTTATAATCCACACCTTTAAACAGCGTGTATCCATCATATCAAT 1014
386 ATAAATTTATTAAAGTTTCATCAATCCAAATCAATGATGTTGAAATCCACCATCCCCA 327
1015 ATTAAAGTTTTTACGCAACGCTTAGATGATGTCACGCTCTACTGAAGGTATATAAGAA 1074
326 GTGAAGTTTACAAAGACGGTTAGATGATATCCACAGTTACTTGTGATGATTTAAGCAC 267
1075 GGTAAAAACAATAATGTTAAGTATGTTGCCAGGTTATAA 1113
266 GCGAGGAATTCGCGAAAAGTTGTTGCGCTCTTGAA 228
RESULT 12
SCCHXIV43/c
LOCUS S.cerevisiae chromosome XIV DNA (43.5 kb). linear PLN 17-FEB-1997
DEFINITION
ACCESSION 246843
VERSION 246843.1 GI:861113
KEYWORDS adenosine deaminase; cyclase-associated protein; FK506-binding
protein proline rotamase; mating pheromone alpha-factor 2;
mitochondrial outer membrane component; mitochondrial ribosomal
protein; NH3 permease; Phe-tRNA; sn-1,2-diacylglycerol
cholinephosphotransferase; tyrosine phosphatase.
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 25310 to 25400)
AUTHORS Bull, P., Thorikay, M., Moenne, A., Wilkens, M., Sanchez, H.,
Valenzuela, P. and Venegas, A.
TITLE The yeast tRNA (Phe) gene family: structures and transcriptional
activities reveal member differences not explained by intragenic
promoters
JOURNAL DNA 6 (4), 353-362 (1987)
MEDLINE 88003973
PUBMED 3308382
REFERENCE 2 (bases 2855 to 2968)
AUTHORS Michaelis, S. and Herskowitz, I.
TITLE The a-factor pheromone of Saccharomyces cerevisiae is essential for
```



```
/codon_start=1
/label=JTB38
/product="mating pheromone a-factor 2"
/protein_id="CAA6681.1"
/db_xref="GI:854493"
/db_xref="SWISS-PROT:P34166"
/translation="MQPITASTQATQKDSSEKKNYIIKGLFWDPAVCVIA"
/complement(3261..5483)
/gene="orf5"
/complement(3261..5483)
/gene="orf5"
/citation=[10]
/codon_start=1
/label=JTD740
/protein_id="CAA6682.1"
/db_xref="GI:854494"
/db_xref="SWISS-PROT:P51907"
/translation="MSSSIFEMTIDHGMSTGLKLSQTASSISMGDEFILCSSTNS
ILUSPLKVFTHIDSITDINTINMEIVSPQGVDDVADKNVLYCIDPYPVEPPCY
DFANPSKVIKPIYEHCRPCLTSVKPESYTPSEVHYTVVSMKMKLYSPFNASRLWN
NFIQINSTQINFIYSIDSLRHKIKYRGDMDFDHSKTSAGDRHHSARSLNAFTT
KSTQPKYDKERICGBIARDEKFLSDEKFLKYSLSQCAKVLPIIDYSRDFVLRN
CEGOPLVQSHVDELIVWAMYLNMGLSLDLLELREMPYVSPRPHPRARFKOH
NKNKNKNSQNSDRNDRSHSLLRSHSTSVVVKVATGNERTPNKRSRSLSLPP
SVDSHSDNSPNDGSLAGESSQTFGLFTSKLRNFKTDSSSKNSMDIGQKWR
SKELNSVQEEIDNNGSTNTNTSVLSFTSPHSLVTAQTSIHENFRSRNSNPIDP
LRCDRSVLKNNFVPEYEGTGRSTISNSAERTNREKLNRENDVNGDEDDDDDA
NYBEDDGVVEDDDINRLVLDRESRYENADVETIIPSTKSFVYTGCLNDFPKK
LSLCKGNQSSNDSKWPATQLVSRKRYIKDSLRCIKPLTEDEHPWVGKLIIFRALPP
```

Query Match
Best Local Similarity 58.8%; Pred. No. 1.7e-69;
Matches 658; Conservative 0; Mismatches 452; Indels 9; Gaps 2;

4 TCAGTTCACACCTCAAAAGCGTCATCATTTGAAGTGACAAAGCTGTGTAAACA 63
24013 TCGATTCCAGAAACCATGAAACCGTGTGTATTGAAAATGCGAAGCGTGTAGTCAACAG 23954

64 GATGCTCAGTTCCAGAAATTAAGAGGGTACAGCGTTGGTGAAGGTTGAGCGTGTGCT 123
23953 GACATTCGATCTGATTAAGAGAGGATTTGTTCTATTAGACTGTGCGGTTGCC 23894

124 GGTAAACCAACTGATGGAAGCATATTGCTTATAGATTGGTCCAGAGGTTCAATCTA 183
23893 GGTAAACCTACCGATTGGAACATATTGATTCAAGATTGGTCTCTCAAGTGCCCTCTTA 23834

184 GGATGTGACATGCTGTGACAGTTGTCAACTTGACCAATGCTAGTACTG----ACTTG 240
23833 GGCTGTGATGACGCGCGCCAAATCGTAAAGTTGGGCCCCAAATGTTGATGCTGCACGCTTT 23774

241 AAGGTTGAGATACCGGTTTCGGTTTGTTCACGCTGCTCCCAACAGATCCTTAAAAAT 300
23773 GCCATTGGTATTACATTTATGGGGTTATTACCGGTTGCTCAGTGAGTTCCTCCCTCAAC 23714

301 GGTGCATTTGCTGAATATGCCAGGTTTATCCACCTTTGTTTATCAA-----GAGTAAC 354
23713 GGTGCTTTGCTGAGTACTCTGCCATTTCATCCGAGACTGCTTATAACAGCCAGAGAG 23654

355 TTAACCTACTCACTGCTGATGAATTTCTGAAGCCCTGTGAAGACTTCGAAATCTGCT 414
23653 TTTAGATTGTGCGTGAAGACAAGCTACCAAGAGCCCGCCGTAATCTTTAGAGAGGGCA 23594

415 GCATCATTTGCGATTTGCTTACCAACTGCTCGTGTGTTGTTGTCATCATCTCGGCTCA 474
23593 GTATCCCTCCAGTCTCATTTACCACCGCTGGTATGATCCTTACACATAGTTTGGGCTG 23534

475 AAAATGGAATGGCACCCTACTACCCGCAACATACTCATCTCATTTATTGATTGGGGTGGT 534
23533 GACATGACATGGAAGCCCTCCAAAGCGCAAGAGATCAACCATCTTATTTTGGGGTGGT 23474

535 GCTACAGAGTGGGTCAACAACTAATCAAGTTCGCAACATATCAATGCTTACTAAG 594
23473 GCCACTGTGTTGGCCAGATGCTTATTCAATTGGCAAAAAAATCAAAACGGTTTCAGCAAG 23414

```
QY 595 ATTGTAACGTGTTGCTTTCTTAAAGCATGAAAGCTTTTAAAGTCTTATGCTGCTGATGAT 654  
Db 23413 ATCATCGTGTGCTTCTCTGTAACATGAAATAATTTGTTGAAGAGTACGCTGCAGATGA 23354

QY 655 GTCTTTGACTATCATGATCAGCGCTTATTGACGACATCAATCGAAGTATCCAAACCTG 714  
Db 23353 CTTTTTGACTTACCACGATGCTGACGTTATTCGACAGATAAAAAAGATACAAACATTT 23294



QY 715 CAACATGTTATTGACGCTGTGGGAAGCATAGTATCCCGAGGCTTATAAAGTCACA 774  
Db 23293 CTTTACTTGGTGGACTGTGCTCCACACAGAAACTATTCAACAGGTGTACAAATGTGCC 23234



QY 775 GCAGATAGTCTACCTCCACATTTATTAGAGTGGTTCCAATGACCACTTTGAAAGCATTCCT 834  
Db 23233 GCTGATGACTTACGCTACGCTCGGTCGTTCAATTGACCGTTTAAACCGAAGATATCAAG 23174



QY 835 GAAGAAATCAGAAAGATTAATGTAATTTGATATTTACTTTGTTGTTATCGTGATCTGGT 894  
Db 23173 GAGNAAGACAGGAGGCAAAACGTCAGTATTGAAAGAACCTTCTTATTTGATAGAGGT 23114



QY 895 CAAGAAATTTCTATTGGGTGCAACAAAGATTTCCTGTAGTCCAGAAATATCATGAAGCCACA 954  
Db 23113 AACGAGTCCATTTGGCAGCTTACTTTCGACAGAGACCTGAAATGATGATGATGATGATG 23054



QY 955 GTTAAATTCGTTAAAGTTTATAATCCACACCTTTAAACAAGCGTGATATCCATCATATGAAT 1014  
Db 23053 ATAAATTTATTAAAGTTTATCAATCCAAATCAATGATGTTGAAATCCACCATCCCA 22994



QY 1015 ATTAAGTTTTTCAGCAACGCTTAGATGATGTCCTGAGCTCTCACTGAGAGTATATAAGAA 1074  
Db 22993 GTCAAGTTTTTACAGAAACGCTTAGATGATATCCACAGTTACTTGTGATATTAAGCAC 22934



QY 1075 GGTAAAAACAAAATCTTAAGTATGTTGCCAGGTATAA 1113  
Db 22933 GGGAGGAATTTCTGCGAAAAGTTGGTTGCGCTTGA 22895



RESULT 13  
SCCHRII/c  
LOCUS 316613 bp DNA linear PLN 29-JUN-2001  
DEFINITION S.cerevisiae chromosome III complete DNA sequence.  
ACCESSION X59720.2 S43845 S49180 S5084 S93798  
VERSION X59720.2 GI:14588895  
KEYWORDS chromosome.  
SOURCE Saccharomyces cerevisiae (baker's yeast)  
ORGANISM Saccharomyces cerevisiae  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Saccharomyces.



REFERENCE 1  
AUTHORS Rad, M.R., Lutzenkirchen, K., Xu, G., Kleinhans, U. and Hollenberg, C.P.  
TITLE The complete sequence of a 11,953 bp fragment from CIG on  
JOURNAL chromosome III encompasses four new open reading frames  
MEDLINE Yeast 7 (5), 533-538 (1991)  
PUBMED 19177317  
REFERENCE 2  
AUTHORS Biteau, N., Frenaux, C., Hebrard, S., Menara, A., Aigle, M. and  
Crouzet, M.  
TITLE The complete sequence of a 10.8kb fragment to the right of the  
JOURNAL chromosome III centromere of Saccharomyces cerevisiae  
MEDLINE Yeast 8 (1), 61-70 (1992)  
PUBMED 92254505  
REFERENCE 3  
AUTHORS Wilson, C., Bergantino, E., Lanfranchi, G., Valle, G., Carignani, G. and  
Frontali, L.  
TITLE A putative serine/threonine protein kinase gene on chromosome III  
JOURNAL of Saccharomyces cerevisiae  
MEDLINE Yeast 8 (1), 71-77 (1992)  
PUBMED 92254506  
REFERENCE 4  
PUBMED 1580103


```


YCL066c, PEP Y, ORF obsolete, is not an ORF in current Chromosome
111 contig.
YCR062w (YCR061w + YCR062w -> YCR061w / C A583 + C B120 -> C B631),
PEP Y, ORF in identical form now C-terminal part of YCR061w, _
YCR068w-a (YCR068w + YCR068w-a -> YCR068w / C C429 + C A145 -> C
A520), PEP Y, ORF in identical form now C-terminal part of YCR068w.
YCR103c (C F11), former ORF in region corresponding to coord.
307255-307587(C) is obsolete, open reading frame is destroyed by
single nt exchange resulting in a
stop codon after 45 aa.
NEW GENETIC ENTITIES:
YCL26c-b (C E193(YCL027c-a)), PEP Y, new ORF
YCL021w-a (C C125), PEP Y, new ORF
ALTERED GENETIC ENTITIES:
YCL076w, PEP Y, peptide elongated N-terminally by 51 aa YCL074w,
PEP Y, 3 aa exchanged
YCL073c, PEP Y, 4 aa exchanged
YCL068c (C F190->C E240), PEP Y, peptide elongated N-terminally by
70 aa YCL065w, PEP Y, 1 aa exchanged
YCL064c, PEP Y, 3 aa exchanged
YCL061c (C D853->C F1096), PEP Y, peptide elongated N-terminally by
31.28; Score 347.8; DB 8; Length 316613;
Query Match
Best Local Similarity 58.3%; Pred. No. 2.5e-67;
Matches 651; Conservative 0; Mismatches 457; Indels 9; Gaps 2;
305474 CAATCCAGAAACCATGAGGCTGCTCATTTGAAGACGGTGAAGCGGTGTCMAAGAGG 305415
5 CAGTTCACACCATCAAAAAGCCGCTCATTTGAAGGTGCAAGCTGTTGTTAAACAG 64
65 ATGCTCAGTTCAGAAATTAAGAGGGGTACAGCTTTGGTGAAGTTGAGCGTGTGCTG 124
305414 CGGTTCCATTCTCGAATTGGAGAGGATTCGTATTGATTAGACACATCGCTGTGCTG 305355
125 GTAACCAACTGATGGAACATATTCCTTATAGATTGTCGAGAGGTTCAATTTAG 184
305354 GTAAACCGAGTGTGTTGGGCACATCTGACTACAGGTCGGGCTCAAGAGATCTATTCTG 305295
185 GATGTGACATTTGCTGTGACAGTGTGCAACTTGGACCA---AATGCTAGTACTGACTTGA 241
305294 GATGTGAGCTGCGGCCAAATGTCAATTTGGCCAGCGCTGATCCTTAAGACTTTT 305235
242 AGGTGAGATACCGGTTTCGGTTTGTCTACGGTGTCTCCCAACAGATCCTAAATG 301
305234 CTATTGTTGATATATTTATGGTTTCATTACGGATCTTCGTAAGTTTCTTCCAAATG 305175
302 GTCATTTGCTGATATGACAGGTTTATCCACCTTTGTTTACAA-----GAGTAACT 355
305174 GTGCTTTGCTGATATTTCTGCTATTTCAACTGTGGTTGCTACAAATCACCCTAATGAAC 305115
356 TAACCTCACTCAACTGCTGATGMAATTTCTGAAGGCCCTGTGAAGAACTTCGAATCTGCTG 415
305114 TCAATTTTGGTGAAGATGTTCTACCTGCGGCCCTGTGACGTTCTTGAAGGGGCAG 305055
416 CATCATGCGAGTTTCGTTGACAACTGCTGGTGTAGTTTGTGTCATCACTTGGGCTCAA 475
305054 CCACTATCCAGTGTCACTGACCAACAGCTGCTGGTGTGTTGACCTATAAATCTGGCTTGA 304995
476 AATGGATGGCCCACTATCCCGGCAACATCACTCATCAATTTGATTTGGGCTGGT 535
304994 ACCTGAATGGAGCCATCAACCCCAAGAAACGCGCCCATCTTATTATGGGCGGCTG 304935
536 CTACAGAGTGGGTCAACAACTAATCAAGTTGCAACATATCAATGCTTTATCAAGA 595
304934 CAATGAGTAGGTGCTGCTCATCCAAATAGCCATAAATGAATGGCTTCAACCAAGA 304875
596 TTGTAACCTGTTCTTCAAAAAGCATGAAAAGCTTTTAAAGTCTTATGCTGCTGATGATG 655
304874 TCATTGTTGCTGCTTCTCGGAAACACGAAAACATGTTGAAGAAATATGCTGCTGATCAAC 304815
656 TCTTTGATATCATGATGAGGCGTTATTGAGCAGATCAAAATCGAAGATATCCAAACCTGC 715
304814 TATTGATATCCATGATATTGACGTGTTAGAACAAATTAACACACAGTACAAACAATATCT 304755

716 AACATGTTATTGACGCTGTGGGAAGCAAGATAGTATCCCGAGGCTCTATAAAGTCACAG 775
304754 CGTATTTAGTCGACTGTGTCGCAATCAAAATAGGCTTCAACAAGTGTACAAATGTGCGG 304695
776 CAGATAGTCTACCTGCCACATTTATAGAAGTGGTTCCTCAATGACCATTTGAAAGCATTCCTG 835
304694 CCGATTAACAGGATGCTACCGTTGTGCAATTTAACTAATTTTGACAGAAGAAACGTCANAA 304635
836 AAGAAATCAGAAAGATAATGTTAAATTTGATATTACTTTGTTGTTGATCGTGCATCTGGTC 895
304634 AGGAGATAGGAGGCAAAATGTCATTTATGACAGAACAAGACTGTTATTCATAGGCGGCC 304575
896 AAGAAATTTCTATTGGGTGCAACAGATTTCTCTCTAGTCCAGAAATATCATGAAGCCACAG 955
304574 ATGAAGTACCATTTGTTGGTACATTTCTCTCTGACCCAGAGCCAGGAGCTGCCA 304515
956 TTAATTCGTTAAGTTTATAAATCCACACCTTTAAACACGGTGTATATCCATCATATGATA 1015
304514 CCGAATTCGTCAGTTCATCAATCCAAAGATTAGTGTATGGGCAATTTCCACATATTCAG 304455
1016 TTAAGTTTTCAGCAACGGCTTAGATGATGTCCTCAGCTCTCACTGAAGGTATAAAGAAG 1075
304454 CAAGGGTCTATAGAAGCGGCTTTACGATGTTCTCTGATCTCTGGAAGACATTTAAATCG 304395
1076 GTAAAAACAAAAATGTTAAAGTATGTTGCCAGGTTTATA 1112
304394 GTAAAGACTCTGCTGAAAGAACTAGTTGCCGCTATTAA 304358
RESULT 14
AX524817
LOCUS AX524817 1122 bp DNA linear PAT 21-NOV-2002
DEFINITION Sequence 7 from Patent EP1236796.
ACCESSION AX524817
VERSION AX524817.1 GI:25169913
KEYWORDS
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE
AUTHORS Yamamoto, H. and Kimoto, N.
TITLE Novel enzyme reduction isolated from Kluyveromyces lactis, methods for producing same, and methods for selectively reducing a carbon-carbon double bond of an alpha, beta-unsaturated ketone using the reduction
JOURNAL Patent: EP 1236796-A 7 04-SEP-2002;
Daicel Chemical Industries, Ltd. (JP)
FEATURES
Location/Qualifiers
1. 1122
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/db_xref="taxon:4932"
7. 1113
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAD57036.1"
/db_xref="GI:25169914"
CDS
/translation="MKAVVIEDGKAVVKEGVPIPELEEGFVLKTLAVAGNPTDWAHI
DYKVGQSGILGDAAGQIVKLGPAVDKPDPSIGDYIYGFHGSVRFPSNGAFABYS
AISTVAVKSPNELKFLGDELVPAGVRLGGAATIPVSLTTAGLVLTYNLGLNKLWE
PSTPONGPILLWGGATVAGOSLIOLANKLNGFTKIIIVASRKHKLKEYGADQLPD
YHIDVVEQIKHYNNISYLDCVANQNTLOQVYKCAADKQDQATVVELTNTLENVKK
ENRRQNTIDRLTSLYSIGHVFFGGITFPADPEARRAATPEPVKFINPKISDQIHHI
PARVYKNGLYDPRILEIDIKIGNSGEKLAVLN"
BASE COUNT 330 a 239 c 255 g 298 t
ORIGIN
Query Match 31.0%; Score 345.2; DB 6; Length 1122;
Best Local Similarity 58.5%; Pred. No. 1.3e-66;
Matches 643; Conservative 0; Mismatches 448; Indels 9; Gaps 2;

```
QY 22 AAAGCCGTCATCATTTGAAGGTGACAAAGCTGTTGTTAAACAGATGTCCTCAGTTCCAGAA 81
DB 10 AAAGCTGTCGTCTATTGAAGACGGTAAAGCGTTGTCAAGAGGCGTTCCTCATCTGAA 69
QY 82 TTAAGAGGGGTACAGCCCTTGGTGAAGGTTGAGGCTGTTCTGGTAAACCAACTGATG 141
DB 70 TTGGAAGAAGGATTCGTATTGATTGAAGACACTCGCTGCTGCTGAACCCGACTGATTGG 129
QY 142 AAGCATATTGCTTATAGATTGGTCCAGAGGTTCAATTTCTAGGATGTGACATGCTG 201
DB 130 GCACACATTTGACTACAAAGGTGGGCGCTCAAGGATCTATTCTGGGATGTGACGTCGGC 189
QY 202 ACAGTTGTCAACTTGACCA---AATGCTAGTACTGACTTGAAGGTTGGAGATACCGGT 258
DB 190 CAAATTGTCAATTTGGGCCAGCGCTGATCCTTAAGACATTTTCTATTGTTGATATATT 249
QY 259 TTGGTTTGTTCACGCTGTTCCAAACAGATCTCTAAAGGTTTGAAGGTTGGAGATACCGGT 318
DB 250 TATGGGTTTCAATTCACGCGATCTTCGTAAGGTTTCTTCCAAATGGTGTCTTTGCTGAAT 309
QY 319 GCAGAGGTTTATCCACCTTTGTTTACAA-----GAGTAACTTAACCTCACTCACTGCT 372
DB 310 TCTGCTATTCAACTGTTGGCTGCTACAAATCAACCAATGAACCTCAAAATTTTGGGTGAA 369
QY 373 GATGAAATTTCTGAAGGCCCTGTGAAGAACTTCGAATCTGCTGCATCATTTGCCAGTTTCG 432
DB 370 GATGTTCTACTCGCGGCCCTGTGACGTCTTTGGAGGGGCGACCACTATCCAGTGCTCA 429
QY 433 TTGACAACTGCTGTTGTTGTTGTTGTCATCACTTTGGGCTCAAAATTTGAATGCGACCCA 492
DB 430 CTGACACACAGCTGGCTGGTGTGACCTATAAATCTGGGCTTGAACCTGAAAGTGGAGCCA 489
QY 493 TCTACCCGCGCAACATCTCATCTCAATTTGATTTGGGGTGGTGTCTACAGCAGTGGTCAA 552
DB 490 TCAACCCGCAAAAGAAACGCGCCCATCTTATTATGGGGCGGTGCAACTGCAGTAGTGAG 549
QY 553 CAACATACTCAAGTTGCAACATATCAATGCTTATTAAGATTTGAATGTTGCTTCT 612
DB 550 TCGTCTATCAATTTAGCAATTAATGATGCTTCAACAGATCATTTGTTGGCTTCT 609
QY 613 AAAAGCATGAAAGCTTTTAAAGTCTTATGGTCTGATGATGCTTTTGATCATATGAT 672
DB 610 CGGAAACACGAAACCTGTTGAAAGATATGGTCTGATCAACTATTGATTACCATGAT 669
QY 673 GCAGGCGTTATTGAGCAGATCAATCGAAGTATCCAACTGCAACATGTTATTGAGCT 732
DB 670 ATTGACGTGTGAGAACAAATTTAAACACAGGTACAAATATCTCGTATTGTCGACTGT 729
QY 733 GTGGGAAGCGAAGATAGTATCCCGAGCGCTTATAAGTCAAGCAGATAGTCTACCTGCC 792
DB 730 GTGCGAATCAAAATACGCTTCAACAGTGTACAAATGTGCGGCGGATTAACAGGATGCT 789
QY 793 ACATATTATGAAGTGGTTCCAAATGACCAATTTGAAAGCATTTCTGAAAGAAATCAGAAAAGAT 852
DB 790 ACCGTTGCGAAATTAACATAATTTGACAGAAAGAAACGTCAAAAGAGAGATAGAGGCAA 849
QY 853 AATGTTAAATTTGATTTACTTTGTTGATCGTGATCTCGTCAAGAAATTTCTATTGGGT 912
DB 850 AATGTCTATTGACAGAACAGACTGTATTCAATAGGCGGCCATGAAGTACCAATTTGGT 909
QY 913 GCAACAGATTTCTGCTGCTAGTCCAGAAATATCATGAAGCCACAGTTAAATTTGTTAGTTT 972
DB 910 GGCATTTACTTTTCCCTGCTGACCCGAGGAGGAGCTGCGACCGAATTTGCTCAAGTTC 969
QY 973 ATAAATCCACCTTAAACAGGTTGATATCCATCATATGAATATTAAGTTTTCAGCAAC 1032
DB 970 ATCAATCCAAAGATTAGTGATGGCAAAATTCACCATATTCCAGCAAGGCTTATAAGAAC 1029
QY 1033 GGCTTAGATGATGTCAGCTCTCACTGAAGGTTATAAAGAGGTTAAACAAATATGTT 1092
DB 1030 GGGCTTTACGATGTTCTCGTATCTCGTATCTCGTAAAGACATTTAAATTCGTAAGAACTCTGGTGA 1089
QY 1093 AAGTATGTTGCCAGGTTATA 1112
```

```
DB 1090 AAACGTTGGCGTATTAAA 1109
RESULT 15
BD174424
LOCUS
DEFINITION
    BD174424
    Novel enone reductase, process for producing the same, and method
    for selectively reducting carbon-carbon double bonds of alpha,
    beta-unsaturated ketone using the same.
    1122 bp DNA linear PAT 18-MAR-2003
ACCESSION
    BD174424
    GI:29120112
VERSION
    JP 2002247987-A/4.
KEYWORDS
    Saccharomyces cerevisiae (baker's yeast)
SOURCE
    Saccharomyces cerevisiae
    Zukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
    Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
ORGANISM
    Yamamoto,H. and Kimoto,K.
    Novel enone reductase, process for producing the same, and method
    for selectively reducting carbon-carbon double bonds of alpha,
    beta-unsaturated ketone using the same
    Patent: JP 2002247987-A 4 03-SEP-2002;
    DAICEL CHEMICAL INDUSTRIES LTD
COMMENT
    OS Saccharomyces cerevisiae (yeast)
    PN JP 2002247987-A/4
    PD 03-SEP-2002
    PF 23-FEB-2001 JP 2001049363
    PI HIROAKI YAMAMOTO, KUNIHITO KIMOTO
    PC C12N15/09, C12N1/19, C12N1/21, C12N5/10, C12N9/02, C12P7/
    26//
    PC (C12N15/09, C12R1:645), C12N15/00, C12N5/00, (C12N15/00, C12R1:645)
    CC Novel enone reductase, process for producing the same, and CC
    method for
    CC selectively reducting carbon-carbon double
    bonds of alpha, beta-
    CC ketone using the same
    CC unsaturated
    FH Key Location/Qualifiers
    FT CDS (7)..(1113).
FEATURES
    source
    1..1122
    /organism="Saccharomyces cerevisiae"
    /mol_type="genomic DNA"
    /db_xref="taxon:4932"
BASE COUNT
    330 a 239 c 255 g 298 t
ORIGIN
    Query Match 31.0%; Score 345.2; DB 6; Length 1122;
    Best Local Similarity 58.5%; Pred. No. 1.3e-66;
    Matches 643; Conservative 0; Mismatches 448; Indels 9; Gaps 2;
QY 22 AAAGCGTTCATCATTTGAAGGTGACAAAGCTGTTGTTAAACAGATGTCCTCAGTTCCAGAA 81
DB 10 AAAGCTGTCGTCTATTGAAGACGGTAAAGCGTTGTCAAGAGGCGTTCCTCATCTGAA 69
QY 82 TTAAGAGGGGTACAGCCCTTGGTGAAGGTTGAGGCTGTTCTGGTAAACCAACTGATG 141
DB 70 TTGGAAGAAGGATTCGTATTGATTGAAGACACTCGCTGCTGCTGAACCCGACTGATTGG 129
QY 142 AAGCATATTGCTTATAGATTGGTCCAGAGGTTCAATTTCTAGGATGTGACATGCTG 201
DB 130 GCACACATTTGACTACAAAGGTGGGCGCTCAAGGATCTATTCTGGGATGTGACGTCGGC 189
QY 202 ACAGTTGTCAAACTTGACCA---AATGCTAGTACTGACTTGAAGGTTGGAGATACCGGT 258
DB 190 CAAATTGTCAATTTGGGCCAGCGCTGATCCTTAAGACATTTTCTATTGTTGATATATT 249
QY 259 TTGGTTTGTTCACGCTGTTCCAAACAGATCTCTAAAGGTTTGAAGGTTGGAGATACCGGT 318
DB 250 TATGGGTTTCAATTCACGCGATCTTCGTAAGGTTTCTTCCAAATGGTGTCTTTGCTGAAT 309
QY 319 GCAGAGGTTTATCCACCTTTGTTTACAA-----GAGTAACTTAACCTCACTCACTGCT 372
```



```
Db 310 TCTGCTATTTCAACTGTGGTTCCTACAAATCAACCAATGAACCTCAAAATTTTGGGTGAA 369
Qy 373 GATGAAATTTCTGAAGCCCTGTGAAGAACTTCGAATCTCTGTCATCATTCGCCAGTTTCG 432
Db 370 GATGTTTCTACTGCGGCCCTGTGAGTCTTTGAGGGGAGCCACTATCCCAAGTGTCA 429
Qy 433 TTGACAACTGCTGTGTAGTTTGTGTATCATCAGCTTGGGCTCAAAAATGGAATGGACCCCA 492
Db 430 CTGACCAACAGCTGCTTGGTGTGACCTATTAACCTTGGCTTGAACCTGAAGTGGAGCCA 489
Qy 493 TCTACCCGGGAACATCATCTCAATTAATGATTTGGGGTGTGTACAGAGTGGGTCAA 552
Db 490 TCAACCCCAACAAAGAAACGGCCCATCTTATTATGGGGCGGTGCAATCGCAGTAGGTGAG 549
Qy 553 CAACATATCAAGTTGCGCAACATATCAATGCTTATTAAGATTTGAATGTTGCTTCT 612
Db 550 TCGCTCATCAATAGCAATAAATGAATGGCTTCCCAAGATCATTTGTTGGCTTCT 609
Qy 613 AAAAAGCATGAAAGCTTTTAAAGTCTTATGGTCTGATGATGTCTTTGACTATCATGAT 672
Db 610 CGGAAACACGAAAGCTTTCGAAAGATATGCTCTGATCAACTATTTGATTACCATGAT 669
Qy 673 GCAGGCGTTATGAGAGATCAAAATCGAAGTATCCAACTGCAACATGTTATTGACGCT 732
Db 670 ATTGACGTGTGAGCAAAATTAACACAAAGTACAAATATCTCTGTTATTGATCGACTGT 729
Qy 733 GTGGGAAGCAAGATAGTATCCCGAGGCTTAAAGTCAAGCATGATGTCTACCTGCC 792
Db 730 GTCCGGAATCAAAATACGCTTCAACAAAGTGTACAAATGTGCGCGCGATTAACAGAGTGT 789
Qy 793 ACATTATTAGAAAGTGTTCATCAATGACCATTTGAAAGCATTTCTGAAAGAAATCAGAAAAAGAT 852
Db 790 ACCGTTGCAATTAATTTGACGAGAAAGAACTCAAAAGGAGATAGAGGCAA 849
Qy 853 AATGTTAAATGATATTAATCTTTGTTGATCGTGATCTGTGTCAGAAATTTCTATTGGGT 912
Db 850 AATGCTACTATTGACAGAACAAAGCTGTTTCAATAGCGGCCCATGAAGTACCATTTGGT 909
Qy 913 GCAACAGATTTCTGCTAGTCCAGATATCATGAGCCACAGTTAAATTCGTTAAGTTT 972
Db 910 GGCATTAATTTCTGCTGACCCAGAGAGAGCTGCCACCAATTTCTGCAAGTTC 969
Qy 973 ATAATCCACACCTTAAACAGCGTGATATCATATGATTAATTAAGTTTTCAGCAAC 1032
Db 970 ATCAATCCAAAGATTAGTGAGGCAATTTCCATATTTCCAGCAAGGCTCTAAGAAC 1029
Qy 1033 GGCTTAGATGATGCCAGCTCTCACTGAAGGTATAAAGAGGTAAAGCAAAATTTT 1092
Db 1030 GGGCTTTACGATGTTCTCGTATCTCGAAGACATTAATCGGTAAAGAACTCTGGTGAA 1089
Qy 1093 AAGTATGTTGCCAGTTATA 1112
Db 1090 AAACGTCGCGGTATTAAA 1109
```

RESULT 16

CNS06FEH 1083 bp DNA linear STS 10-JAN-2001
LOCUS T7 end of clone AR0AA029E02 of library AR0AA from strain CBS 732 of
DEFINITION Zygosaccharomyces rouxii, sequence tagged site.

ACCESSION AL396303

VERSION AL396303.1 GI:12148292

KEYWORDS STS.

SOURCE Zygosaccharomyces rouxii

ORGANISM Zygosaccharomyces rouxii

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; Saccharomycetaceae; Zygosaccharomycetes;

1 (bases 1 to 1083)

REFERENCE Souciet, J.L., Aigle, M., Artiguenave, P., Blandin, G.,

Boletín-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,

de-Montigny, J., Dujon, B., Durand, P., Lepingle, A., Llorente, B.,

Malpertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S.,

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

Saurin, W., Tekala, P., Toffano-Nioche, C., Wesolowski-Louvel, M.,
Wincker, P. and Weissenbach, J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)
20584711
11152876
2 (bases 1 to 1083)
de Montigny, J., Straub, M., Potier, S., Tekala, P., Dujon, B.,
Wincker, P., Artiguenave, P. and Souciet, J.
Genomic exploration of the hemiascomycetous yeasts: 8.
Zygosaccharomyces rouxii
FEBS Lett. 487 (1), 52-55 (2000)
20584718
11152883
3 (bases 1 to 1083)
Genoscope.
Direct Submission
Submitted (06-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This STS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

COMMENT

yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES

source

1..1083
/organism="Zygosaccharomyces rouxii"
/mol_type="genomic DNA"
/strain="CBS 732"
/db_xref="taxon:4956"
/clone_lib="AR0AA029E02"
/clone_lib="AR0AA"
/notes="end : T7"

misc_feature

<10..>1008
/notes="similar to Saccharomyces cerevisiae ORF YLR460c [
similarity to C.carbonum toxD protein]
similar to Saccharomyces cerevisiae ORF YNL134c [
similarity to C.carbonum toxD gene]"
/evidence=not_experimental

misc_feature

<34..>1008
/notes="similar to Saccharomyces cerevisiae ORF YCR102c [
similarity to C.carbonum toxD gene]"
/evidence=not_experimental

BASE COUNT 330 a 206 c 239 g 306 t 2 others

ORIGIN

Query Match 22.7%; Score 252.4; DB 11; Length 1083;
Best Local Similarity 55.7%; Pred. No. 6.7e-46;
Matches 596; Conservative 2; Mismatches 443; Indels 29; Gaps 5;

Qy 1 ATGTGAGTTCCAACTCAAAAGCGCTCATATTGAAGCTGACAAAGCTGTGTTAAA 60
Db 16 ATGTCAATTCCAACCTAGTATGAAGCTGCTGCTATTTGAAGCGGTAAACGCTGCTGAAG 75
Qy 61 ACAGATGCTCTCAGTTCCAGAAATTAAGGAGGCTACAGCCCTTGGTGAAGGTTGAGGCTGTT 120
Db 76 GAAAGCTTCTTTTACCACCTCTCAGGAGGCTCAAGTGTTCAGACGCTTGCAGCT 135
Qy 121 GCTGGTAAACCAACTGATTGGAGCATATTGTTTATAAGATTGGTCCAGAGGTTCAATT 180
Db 136 GCTGTTAATCCAACGATTCGAGACATTTGAGTATATAAACTTGTGCCAAGCTCAATT 195
Qy 181 CTAGGATGACATTTGCTGTGACAGTTGTCAAACCTTGGACCAAAATGCTAGTACTGACTTG 240
Db 196 TCAGGATGTGACGCTGCGGACAAAGTTGTAAAGTTAGGGGCTAACCGTTAAATAATGTCGCT 255

		755	AATTGATCAAATTAAGACCAAGTACA	CTGAATTTGGTTATTAGACTGTGTC	CCAC	696
Qy		741	CGAAGATAGTATCCCAGGCCTATAAAGTCACAGCAGATAGTCTACCTGCCACATTATT		800	
Dd		695	TCTAGACACTTTCAACCAGTGTAACCAATGTGCTCCCACTGATCGTAAATCAACTGTGCAT		636	
Qy		801	AGAAGTGGTTCCCAATGACCAATTGAAGCATTCCTGAAGAAATCAGAAAAAGATA---	-ATGT	857	
Dd		635	CAATTAATATGGGGTTTACTCTAGATGCTATTAAACCAAGNATTAAGACACA	CAGAAAGATGT	576	
Qy		858	TAAATTTGATATTACTTTTGTGTATCTGTGCATCTGGTCAAGAAATTCATTGGGTGCAAC		917	
Dd		575	CTCCTTCGAGGACACTGCTCTTTACACTTGTTTAGGTTTTTGATGTCATKTTAGTGAATA		516	
Qy		918	AAGATTTCCTCTAGTCCAGNAATATCATGAAGCCACAGTTAAATTCGTTAAGTTTATAAA		977	
Dd		515	CACAGTCAAAGCTGANCCAGANACAGACAGGNANNCATBAAAANATGTGAATTTAATTAA		456	
Qy		978	TCCACACCTTAAACAACGGTGATATCCAATCATATGAATATTAAAGTTTTTCAGCAACGGCTT		1037	
Dd		455	TGATAAATTGACCAATGGTTTCATTGGAACCAATATCCCAATTTAAGGTGTACAAAATGGGGTT		396	
Qy		1038	AGATGATGTCCCCAGCTCTCACTGAAGGTATAAAGAGAAAGGTAACCAAAATGTTAAGTA		1097	
Dd		395	AGAAATCCATTCAAATCATAAATGATTTGGAANAATGGTAANAACCTCGTGTGAAAAAATT		336	
Qy		1098	TGTTGCCAGGTT		1109	
Dd		335	TGTGGCCACTKT		324	

RESULT 18	
YSCDC91A/c	
LOCUS	YSCDC91A
DEFINITION	Saccharomyces cerevisiae cdc9l gene, complete cds.
ACCESSION	L31649
VERSION	L31649.1 GI:469465
KEYWORDS	.
SOURCE	Saccharomyces cerevisiae (baker's yeast)
ORGANISM	Saccharomyces cerevisiae
	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

BASE COUNT	1621 a	975 c	933 g	1836 t
ORIGIN				
Query Match		9.3†	Score 103.2;	DB 8; Length 5355;
Best Local Similarity		54.1†	Pred. NO. 1.2e-12;	

Matches	210;	Conservative	0;	Mismatches	178;	Indels	0;	Gaps	0;
Qy	725	TTGACGCTGTGGGAACGAGATAGTAGTATCCCGAGGCCTATAAAGTCACAGCAGATAGTC	784						
Db	5354	TCGACTGTGTGCGGAATCAAGATACGCTTCAACAAGTGTACAAATGTCGCGCCGATAAAC	5395						
Qy	785	TACCTGCCACATTATTAGAAAGTGGTTCGAATGACCAATTGAAAGCATTCTCTGAAGAATCA	844						
Db	5294	AGGATGCTACAAATTGTTGAAATTAATAAATTTGACAGAAGAAAACGTCAAAAAGAGAACA	5235						
Qy	845	GAAGAATAATGTTAAAAATTGATATTACTTTGTTGTATCGTGCATCTGGTCAAGAAATTC	904						
Db	5234	GGAGACAACAGTTACTATTGACAAATAAAGGCTATTCATATAGTGGCCATGAAGTAC	5175						
Qy	905	TATTGGGTGCACCAAGATTTCCTGTGATGCCAGATATCATGAAGCCACAGTTAAATTCC	964						
Db	5174	CAITTCGAAACATTACTTTACACGCCACTCAGAAGCTAGGAAAGCTGCAATAAAATTTA	5115						
Qy	965	TTAAGTTTATAATCCACACCTTTAAACAACGGTGATATCCATCATATGATATATAAGTTT	1024						
Db	5114	TCAAAATTCATCAATCCAAAGATTAAATGATGGACAAATTCGCCATATATCCAGTAAGGTCT	5055						
Qy	1025	TCAGCAACGGCTTAGATGATGTCCAGCTCTCACTGAAGTATATAAAGAAGGTAAAAACA	1084						
Db	5054	ATAAGAACGGCTTTGTGATGTTCCATATCTCTAAAGACATCAAAATATGGTAAGAACT	4995						
Qy	1085	AAAATGTTAAGTATGTTGCCAGGTATA	1112						
Db	4994	CTGGTGAAAAACTCGTTGCGTATTAAA	4967						
RESULT 19									
LOCUS	AX536956		1086 bp	DNA	linear				PAT 22-NOV-2002
DEFINITION	Sequence 557 from Patent WO02064766.								
ACCESSION	AX536956								
VERSION	AX536956.1		GI:25263391						
KEYWORDS									
SOURCE									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
FEATURES									
source									
BASE COUNT									
ORIGIN									
Query Match			6.4%;	Score 71;	DB 6;	Length 1086;			
Best Local Similarity			49.0%;	Pred. No. 2e-05;					
Matches	324;	Conservative	0;	Mismatches	310;	Indels	27;	Gaps	4;
Qy	35	TTGAAGGTGCACAAAGCTGTTGTTTAAACACAGATGCTCTCAGTTCAGAAATTAAGAGAGGTA	94						
Db	29	TCGACCTTATCAATTAGCGAAATTAAGATATTCTCAACAAAAATAAAGAAAAATG	88						
Qy	95	CAGCTTGGTGAAGGTTGAGGCTGTTGCTGGTAACCCAACTGATTGGAAGCATATTGCTTT	154						
Db	89	AAATATTAAATCAAAAGCAGTAGCTTTTGCAATAAACCCAACTGATTGGAACCATATTGTTT	148						
Qy	155	ATAAGATTGCTCCAGAGGTTTCAATTCTAGGATGTGACATTGCTGGTACAGTTGTCAAC	214						
Db	149	ATCAATTTGGCAGCCCGCAGGTGATGTTGTTGGTTCGATGTTAGTGGATCAATTGAAGAAG	208						
Qy	215	TTGACCCAAATGCTAGTACTGACTTCGAAGGTTGGAGATACCGGTTTTCGGTTTTGTTCAACG	274						

Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hoques, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowals, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhera, L., Loulseghe, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milobavljovic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokelameh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, P., Poindexter, A., Popovic, D., Primus, E., Pu, L., Fu, L., Fu, L., Fu, L., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojase, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorrelle, R., Soosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission

Unpublished

2 (bases 1 to 219955)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (07-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 219955)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On May 10, 2003 this sequence version replaced gi:25188698. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GHKS
Center clone name: CH230-141C19
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 200763 bases at least Q40
Consensus quality: 203589 bases at least Q30
Consensus quality: 205400 bases at least Q20

Estimated insert size: 210228; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 2 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
* 1 158013: contig of 158013 bp in length
* 158014 158113: gap of unknown length
* 158114 212955: contig of 54842 bp in length.

FEATURES
source
1. 212955
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-141C19"
3356. 4203
/note="clone boundary
clone end:T7
site:EcoRI
end sequence:RWAXI22TJC"
3361. 4184
/note="clone boundary
clone end:T7
site:EcoRI
end sequence:RWAXI22TJB"
39881. 41507
/note="wgs end_extension
clone end:T7"
71654. 73136
/note="wgs end_extension
clone end:T7"
158114. 159255
/note="wgs end_extension
clone end:T7"
159306. 161347
/note="wgs end_extension
clone end:T7"

misc_feature
54541 a 47359 c 49948 g 54864 t 6243 others
ORIGIN

Query Match 4.4%; Score 48.6; DB 2; Length 212955;
Best Local Similarity 47.5%; Pred. No. 1.5;
Matches 144; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

QY 25 GCGTCATCATTCGAGGTGACAAAGCTGTGTAAACAGATGCTCAGTTCAGATTA 84
Db 156322 GCTGCTGTTATTGCGATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTTT 156263
QY 85 AAGAGGGGTACAGCCTTGGTGAAGGTTGAGGCTGTGGTAAACCACTGATTCGAG 144
Db 156262 GTTGATGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 156203
QY 145 CATATTGCTTATAAGATTGGTCCAGAGGTTCAATTCTAGGATGTGACATTCGTG 204
Db 156202 CTTGCTGCTGCTTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 156143
QY 205 GTTGTCAAACTGGACCAATGCTAGTACTGACTTGAAGTTGGAGATACCGGTTCC 264
Db 156142 GATGTTGCTGCTATTCCTGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 156083
QY 265 TTGTTTCCAGGTGCTTCCCAACAGATCCTAAAAATGGTGCATTTGCTGAATGCG 324
Db 156082 GTTGTTCTTGATGATGATGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 156023
QY 325 GTT 327
|||


```

/gene="At2g44525"
/notes="synonym: F16B22.47; supported by full length cDNA:
Ceres:8402"
mRNA
join(2606..2812,2969..3085,3370..3404,3546..3646,
3744..4019)
/gene="At2g44525"
join(2640..2812,2969..3085,3370..3404,3546..3646,
3744..3830)
/gene="At2g44525"
/codon_start=1
/product="expressed protein"
/protein_id="AAM14882.1"
/db_xref="GI:20197026"
LIDNPDDQLRFOENFTSFTVNGKYEGSLLCVGNLMSWSPKSEITFTDLSLIFO
TVRPIPELLVGGORDHPVTPVQRQPVKSLGKMLTVDNRNAASTYINILNBSGRVVA
AALLPVGVTS"
4242..6380
/gene="At2g44530"
/notes="synonym: F16B22.2"
join(4242..4337,4472..4648,4876..5124,5395..5538,
5646..5813,5885..5954,6035..6116,6206..6380)
/gene="At2g44530"
join(4242..4337,4472..4648,4876..5124,5395..5538,
5646..5813,5885..5954,6035..6116,6206..6380)
/gene="At2g44530"
/codon_start=1
/product="putative phosphoribosyl pyrophosphate
synthetase"
/protein_id="AAC27455.1"
/db_xref="GI:3341673"
/translacion="MASIVQSPSTFPALNLRSLIRPPSVRFPLKCNADPYKFDG
GNAGPHLLGDTVPAGESRTRLEDSTYQNTLRIFSGTANPLAQEISCVLGLDGL
KIKRPADEGIVQLQESVRCGDVFLVQTPCPANENMELLMIDACRAAKTIT
AVIPFGADAKRTQRESIAKLANLITQSGADRVLACDLHSGMSGYFDIPVDH
VYQGPVLDLAKVAIIVDDMIDTAGTISKGAALLHOEGAREYVACTTHAVFPFPAIS
RLSGLFQEVITNTIPLSEKPPQTLVSLVANLIGETIWRVHDDCSVKT"
complement(6742..6763)
/rpt_family="AT_rich"
/complement(6851..8620)
/gene="At2g44540"
/notes="synonym: F16B22.3"
complement(join(6851..7168,7280..7561,7648..8301,
8399..8620))
/gene="At2g44540"
complement(join(6851..7168,7280..7561,7648..8301,
8399..8620))
/gene="At2g44540"
/codon_start=1
/product="putative cellulase"
/protein_id="AAC27456.1"
/db_xref="GI:3341674"
/translacion="MSSQKNGSSQCLWTSICIVLVMSMARGAVSTNYAELQKSLLY
FEAORSKGLPNQRTWRGSDALRDGSDAHVDLTGGYVDAGDNMKFGPPLAFTTTLA
WSNIEMSELRAHEQCNRLKAWTDYLIKAHPQPNVLYQGVGEGNSDHCWMPRE
DMTPRTSYRIDAQHPGSLAGETAAMAAASIAFAPDKAYANILIGHAKDLFAFAK
AHGGLYQNSIPNAGGFYASSGYEDLLWAAWLHRTATDQIYLDYLTQASNGGPRSV
FAMDKFLVGAQVLVAKLVEGVKQEGKMIYKSMAEQFICNCAQKGNVKTGGGL
LYFLPWNLYQYTTATFVLSAYSKYLTKAKASIQCPKQALQASDLQLQARSQVDYILG
SNPQKSMYGVGTNYPKPHRAASIVSRKDKTPVTCSGGYDKWYNNPAPNPVLM
GALVGGPNNDVYGDERSNQAEPAATVTPVAVGVLAAPV"
complement(9891..11652)
/gene="At2g44550"
/notes="synonym: F16B22.4"
complement(join(9891..10208,10319..10600,10686..11336,
11431..11652))
/gene="At2g44550"
complement(join(9891..10208,10319..10600,10686..11336,
11431..11652))
/gene="At2g44550"
/codon_start=1
/product="putative glucanase"

```

```

/protein_id="AAC27457.1"
/db_xref="GI:3341675"
/translacion="MSQLKNGSSQCLWTSICIVLVMSMARGAVSTNYAELQKSLLY
FEAORSKGLPNQRTWRGSDALRDGSDAHVDLTGGYVDAGDNMKFGPPLAFTTTLA
WSNIEMSELRAHEQCNRLKAWTDYLIKAHPQPNVLYQGVGEGNSDHCWMPRE
DMTPRTSYRIDAQHPGSLAGETAAMAAASIAFAPDKAYANILIGHAKDLFAFAK
AHGGLYQNSIPNAGGFYASSGYEDLLWAAWLHRTATDQIYLDYLTQASNGGPRSV
FAMDKFLVGAQVLVAKLVEGVKQEGKMIYKSMAEQFICNCAQKGNVKTGGGL
LYFLPWNLYQYTTATFVLSAYSKYLTKAKASIQCPKQALQASDLQLQARSQVDYILG
SNPQKSMYGVGTNYPKPHRAASIVSRKDKTPVTCSGGYDKWYNNPAPNPVLM
GALVGGPNNDVYGDERTDYQHAEPAPATAAPFVGVLAAPV"
complement(14607..14630)
/rpt_family="AT_rich"
/complement(14693..14723)
/rpt_family="AT_rich"
complement(14935..16778)
/gene="At2g44570"
/notes="synonym: F16B22.6"
complement(join(14935..15252,15445..15729,15807..16460,
16557..16778))
/gene="At2g44570"
/codon_start=1
/product="putative glucanase"
/protein_id="AAC27459.1"
/db_xref="GI:3341677"
/translacion="MSCISQCFITIKSICIVLLLSITCGAVSANSYGEALTKSLLY
FEAORSKGLPNQRTWRGSDALRDGSDAHVDLTGGYVDAGDNMKFGPPLAFTTTLA
WSNIEMSELRAHEQCNRLKAWTDYLIKAHPQPNVLYQGVGEGNSDHCWMPRE
DMTPRTSYRIDAQHPGSLAGETAAMAAASIAFAPDKAYANILIGHAKDLFAFAK
DYPGVYQNSIPNAGGFYASSGYEDLLWAAWLHRTATDQIYLDYLTQASNGGPRSV
FAMDKFLVGAQVLVAKLVEGVKQEGKMIYKSMAEQFICNCAQKGNVKTGGGL
LYFLPWNLYQYTTATFVLSAYSKYLTKAKASIQCPKQALQASDLQLQARSQVDYIL
GSPKQNSYGVGTNYPKPHRAASIVSRKDKTPVTCSGGYDKWYNNPAPNPVLM
VGAIVGGPEPDYAGDERSDFQHGEPDVTVPAPLVGLAAPV"
complement(16815..16914)
/rpt_family="(TA)n"
complement(17216..17237)
/rpt_family="AT_rich"
complement(18506..23922)
/gene="At2g44580"

```

Query Match 4.3%; Score 48.2; DB 8; Length 109375;
Best Local Similarity 50.7%; Pred. No. 1.9;
Matches 116; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

Qy 800 TAGAAGTGGTTCCAAATGACCATTTGAAGCATTCCTCGAAGAAATCAGAAAGATAATGTTA 859
|||||
Db 19728 TAGTATATGAAACAACACTAAATTTGTCGACATTATTGCTTGTGATTTGTTAAAGTTA 19787
|||||

```
Qy 860 AAATGATATTAATCTTTGTTGATCGTGCATCTGGTCAAGAAATCTATTGGTGCAACAA 919
Db 19788 AAAATTAATTAATGTTTATTAATTTGTAAGCTCGTTTACATTAATAGTTGGTACACAA 19847

Qy 920 GATTTCCTGCTAGTCCAGAAATATCATGAAGCCACAGTTTAAATTCGTTAAGTTTATAATC 979
Db 19848 TCTCTCCCTATGAACAATAATAAAACAATAACAATAAAACATCAAAAATGAATTAAGTC 19907

Qy 980 CACACCTTAACACGGTGATATCATCATATGAATATTAAGTTTTCAG 1028
Db 19908 ATTACTTTCCGAAGTTTACTGTCAGTTTAAACAATAACGTAGTCCAG 19956

RESULT 24
PFMAL1P2_3
WPCOMMENT
Sequence split into 4 fragments LOCUS PFMAL1P2 Accession AL031745
Fragment Name Begin End
PFMAL1P2_0 1 110000
PFMAL1P2_1 100001 210000
PFMAL1P2_2 200001 310000
PFMAL1P2_3 300001 384550
Continuation (4 of 4) of PFMAL1P2 from base 300001 (AL031745 Plasmodium falciparum DNA)

Query Match 4.3%; Score 47.6; DB 3; Length 84550;
Best Local Similarity 43.2%; Pred. No. 2.7;
Matches 227; Conservative 0; Mismatches 299; Indels 0; Gaps 0;

Qy 563 AAGTTGCGAAACATATCAATCTTATCTACTAAGATTTGAACCTGTGCTTCTTAAAGCATG 622
Db 32159 AATGTGATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 32218

Qy 623 AAAAGCTTTTAAAGCTTATGCTGCTGATGATGCTTTGACTATCATGATGCGCGTTA 682
Db 32219 ATAATTTTGATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 32278

Qy 683 TTGAGCAGATCAATCGAAGTATCAAACTGCAACATGTTATTGACGCTGTGGGAAGCG 742
Db 32279 GTGATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 32338

Qy 743 AAGATGATATCCCGAGCGCTTAAAGTCAAGCAGATAGTCTACCTGCCACATTAATAG 802
Db 32339 GTCATAATATTTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 32398

Qy 803 AAGTGGTTCATGACCATGCAAGCATTCCTGAAGAAATCAGAAAGAGATAAGTTTAAA 862
Db 32399 ATAAACAAGAAAGATGAATGATCTATCAGATGACAAAGATGGAAGAGAGAGAGAGAG 32458

Qy 863 TTGATATTACTTTGTTGATCGTCATCTGCTCAAGAAATCTTATTGGTGCAACAGAT 922
Db 32459 TAGATAATGTTTAACTTAATCTGTTAGTATGATAACAATATATGATGATGATAATA 32518

Qy 923 TTCCTGCTAGTCCAGAAATATCATGAAGCCACAGTTAAATTCGTTAAGTTTATAAATCC 982
Db 32519 GTAGTATGATAATAATAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 32578

Qy 983 ACCTTAACACGGTGATATCATCATATGAATATTAAGTTTTCAGCAAGCGCTTAGATG 1042
Db 32579 ATGATAATATATATAGTATGATGATAATAATAATAATAATAATAATAATAATAATAATA 32638

Qy 1043 ATGTCCCGAGCTCTCACTGAAGGTATAAAGAGGTATAAAGAGGTATAAAGAGGTATA 1088
Db 32639 ATATATATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 32684

RESULT 25
AE007561
LOCUS 13844 bp DNA linear BCT 27-JUL-2001
DEFINITION Clostridium acetobutylicum ATCC824 section 49 of 356 of the complete genome.
ACCESSION AE007561 AB001437
VERSION AE007561.1 GI:15023317
```

KEYWORDS

SOURCE ORGANISM

Clostridium acetobutylicum
Clostridium acetobutylicum
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.

REFERENCE AUTHORS

1 (bases 1 to 13844)
Nolling, J., Breton, G., Omelchenko, M.V., Markarova, K.S., Zeng, Q.,
Gibson, R., Lee, H.M., Dubois, J., Qiu, D., Hitti, J., Wolf, Y.I.,
Tatusov, R.L., Sabathe, F., Doucette-Stamm, L., Soucaille, P.,
Daly, M.J., Bennett, G.N., Koonin, E.V. and Smith, D.R.
Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum
J. Bacteriol. 183 (16), 4823-4838 (2001)

TITLE

1 (bases 1 to 13844)
11466286

JOURNAL MEDLINE

21359325

PUBMED

11466286

REFERENCE

2 (bases 1 to 13844)

AUTHORS

Childress, D., Zeng, Q. and Smith, D.R.

TITLE

Direct Submission

JOURNAL

Submitted (24-JUL-2001) GTC Sequencing Center Production,
Finishing, and Bioinformatics teams, Genome Therapeutics Corp., 100
Beaver Street, Waltham, MA 02453-8443, USA

FEATURES

Source

1. .13844
/organism="Clostridium acetobutylicum"

/mol_type="genomic DNA"

/strain="ATCC 824"

/db_xref="ATCC:824"

/db_xref="taxon:1488"

135. .1466

/gene="CAC0460"

135. .1466

/gene="CAC0460"

/codon_start=1

/transl_table=11

/product="CBS-domain containing protein, YHDP B. subtilis
ortholog"

/protein_id="AAK78440.1"

/db_xref="GI:15023318"

/translation="MEADPDPMILQVLIAVLIVNAFFSAEWSIVSLNTRMKI
LSEGRNQAILEKLEKFTLSVQGITLAFSSAFATGSDMFGGLKINI
PYSSQALIVITVLSYFTVFGELFPKRIALQKSEKIAMLAARILYVSKITVPVK
LISASVNLIRITGMNMNLEKSEIIEISIVEAGQEVINESEKIMIDIFPDD
RIAEVMTPTREVLINIDTPLNKVDELLEIKYSRVPVFGSDNIIGILYKDFIV
EARKYFENVNINIIHSPYFPERKNINELFKELQAAKHMALVIDEGGSGIVTI
EDLIEVMGNISDEYEDDNYIKKIDESTFIKGIPIPELNEQLDLDETDYDT
LGLLKLGLHHPEDNEQSIKYKGIITFEIKVEKRIEKEVHLQKDLLE"

1493. .1891

/gene="CAC0461"

1493. .1891

/gene="CAC0461"

/codon_start=1

/transl_table=11

/product="Mercuric resistance operon regulatory protein,
MerR family"

/protein_id="AAK78441.1"

/db_xref="GI:15023319"

/translation="MRYIKIGELSKIVNLISITIRYRKIRLIPKPERNNRYKKYSE
MTYIRKIIKAKNYGFTLHEISNLSFSEVSTSIEDIDLSSTIMMKVQIIECEKRLN
YKKNQLKPFEEINLQCPVLNRIINEKNT"

1955. .3151

/gene="CAC0462"

1955. .3151

/gene="CAC0462"

/codon_start=1

/transl_table=11

/product="Protein of short-chain alcohol dehydrogenase
family"

/protein_id="AAK78442.1"

/db_xref="GI:15023320"

/translation="MIVRAKFKVGFIRVDHPYGCRRREVLNQIDCKKALGFRGPKKVL
IVGASSGFLATRIISVAFGGPEAHTIGVSYETGDRDRIGTAGWNNIIFKFKAKKGG
LVAKNFIEDAFSNETKVIKIDFGKIDLVYSLAAPRKDTKGNVYSRIKTI
LGDFRGPTIDVERDITLTKVSSAIEETETRKVMGQEDMQEWCELLYEDCFSDKA

RESULT 25

AE007561

LOCUS

DEFINITION Clostridium acetobutylicum ATCC824 section 49 of 356 of the

complete genome.

ACCESSION AE007561 AB001437

VERSION AE007561.1 GI:15023317


```
TTIAYVIGSPRTYKIYREGTIGIAKQDLEDAKALINEKLNRVITGGRAFVSVNKAALT
KASAYITPFPYLAAILYWKWKENIHENICIMFSEKIIYSNEKIQFDDKGLRMD
DLERKQVQDSEVDRIWSNITPENFELS DYGYKKEFWNLNGFLDGDVYSKDIDIEL
LRKLEP"
/genes="CAC0463"
/complement (3180. .4226)
/genes="CAC0463"
/complement (3180. .4226)
/genes="CAC0463"
/codon_start=1
/transl_table=11
/product="Serine protease Do (heat-shock protein)"
/protein_id="AAK78443.1"
/db_xref="GI:15023321"
/translation="MKTHKLSIVLASVLLTSFTGCTFASNKKKITPDTANAQAST
SVESPIKASKVLPSPVGVVTTTIERDNTKKGVGSMIIDSQGYLLTNHVGAMTS
KDLVSLYDGSIGAKPLWANESLDLSIIKIDKNLOAVTLGDSKKVDIGETAIGN
PLGINFORTSGIVSAVNRVTEAGEVFMEHLLQTDASINPNSGGLPDANGNVIG
VNSAKISAEIGFAVINIVKPVLSKLTITGQFKTPIVIGIGLDSKMGVNLNFKF
GIYVINISPNSGAAAGINKGDIILSVNGRKNINTMNBRESIYTIGANNTVSLKLTGA
SGEKTNVNKIQSST"
/genes="CAC0464"
/4351. .4692
/4351. .4692
/genes="CAC0464"
/codon_start=1
/transl_table=11
/product="Protein, related to divalent cations, tolerance
CutA family"
/protein_id="AAK78444.1"
/db_xref="GI:15023322"
/translation="MQINQKIBIFIPKEYVEKLGDKLSEANVGKIGNYDHCWATNIV
EGYWRPLEGSEPVGEIGKVCHESECKEYVENALKIIRKVHYPEEPVINI
PILNQLFENNR"
/complement (4841. .5632)
/genes="CAC0465"
/complement (4841. .5632)
/genes="CAC0465"
/codon_start=1
/transl_table=11
/product="Predicted transcriptional regulator,
dica/hlpB/ansr family"
/protein_id="AAK78445.1"
/db_xref="GI:15023323"
/translation="WQDSNLEIKILKNKIYDTSIIFTHNLKLIQLEGTKKLAKSI
GVSDLLASKYSGDAPFAETILYICKLYNTIDFELNIPLTSSYFEDMLYDDTGPD
IFKODYFCYRVTVNSGKSGJHESNLRTKNCIERIYIKGENVKEFKGTYKISDNLV
PFNLYSREYGAYITMINKPSLNKSKYIGGIAMLSLPSDANSKPCQCKIIISSKINRE
LSPDILNNFLCFKBSRGSFGNIKISRYEDEKVDYFITNLNEETKL"
/5832. .6254
/5832. .6254
/genes="CAC0466"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK78446.1"
/db_xref="GI:15023324"
/translation="MNLKRVYDIKSLCEYEFQDKIYFQDKIYGVYIYKNSNEGK
YIEFWIITDFTGEINLLKIHDKLFNSKRGYVSLDDVGVKVDLQMKPYYIVEN
GKLKKGYNNIDYFTWMLVKNKAVGKISPSIDSLKGEF"
/6297. .7106
/6297. .7106
/genes="CAC0467"
/codon_start=1
/transl_table=11
/product="Uncharacterized membrane protein, homolog of
yDAH B subtilis"
/protein_id="AAK78447.1"
/db_xref="GI:15023325"
/translation="MNIQILLILFNFIITPITGLAYSTRLVGVTKIAISFTVFN
LTLVSRVATVFBPLVYAEKNLYSSNLIVFNSIILVSGAASILGAVLIPTRMF
SKGVNFFSLEKSLPRLVHSMTRRGIVSMQCVVLPSRKSVEKENYRKLPRRIIVYNF
```

```
ISVAALTVALAPIYAFKIVNIRATCVTLSSVVNGIAQIQLMTIFIDPQMSINTBEVM
ERKCSDEDFKNVCVLANVGVSKVAGTFAAIFILIPASFAIVFAKVGVDFISRI"
/7119. .7823
/genes="CAC0468"
/7119. .7823
/genes="CAC0468"
/codon_start=1
/transl_table=11
/product="HAD superfamily hydrolase"
/protein_id="AAK78448.1"
/db_xref="GI:15023326"
/translation="MIKALIMDSGKVLNKPATGEWIFPKFPFYVDWHERNTIPLIKR
KSAFDKANKYIEQKLIITTEBEEYKCFLEYIKLSKELYQLKNNIDIBETEDLVYN
KPYDYDAIYIEIPKLSKYKLAIVSDAPVLEGVFKXAGLRKYFSSPVISSVKGVT
KPNEMYKTALSELRLVLPBEAVFIDDNINRCEGAVKLGKISFVLSRKSRIYAYNKLFN
RRTKSRNLKQLEEIY"
/8009. .8584
/genes="CAC0469"
/8009. .8584
/notes="spmA"
/genes="CAC0469"
/codon_start=1
/transl_table=11
/product="Spore maturation protein A (gene spmA)"
/protein_id="AAK78449.1"
/db_xref="GI:15023327"
/translation="MINYIWFIFIFVGTIVGLSNGSTLSSSIINSTSSVKLMISL
LGIMCLWCGVMRIAESKGLTDKLSIVLRPLKIVPKDISKNNVMGPMIMNITSNMLG
LGNATPFGIKTMEENQKINKKSTATNDMAKPLVINAAICQLLPTTIISIRACGSS
NPAIIVIPITVSVTVAIMGLIYCKILQRHP"
/8596. .9117
/genes="CAC0470"
/8596. .9117
/notes="spmB"
/genes="CAC0470"
/codon_start=1
/transl_table=11
/product="Spore maturation protein B (gene spmB)"
/protein_id="AAK78450.1"
/db_xref="GI:15023328"
/translation="MNYIISALIPILIFISIVLYGMIKKVKYECFVEGAKDGIRICVN
IPFYIAMLLSVAIFRESGAMQYLINILKPVFKIIGFPSELVPLIIVKPLSGSGALGV
FABILKKYGADTYIGIINSIIMSGSTETIFITYITVYVYGAIKIKIRHTLWAAINADMTS
IIVAIIVVRPMGI"
/9426. .9998
/genes="CAC0471"
/9426. .9998
/genes="CAC0471"
/codon_start=1
/Query Match 4.3%; Score 47.4; DB 1; Length 13844;
Best Local Similarity 52.8%; Pred. No. 3.3;
Matches 102; Conservative 0; Mismatches 91; Indels 0; Gaps 0;
Qy 809 TTCAATGACCATTTGAAAGCATTTCCCTGAAGAAATCGAAGAAATGATAAATTGTTAAATGATA 868
Db 2163 TTGGAGTATCCTATGAACAGGAGCTACAGATAGAGATAGGAACAGCGGATGATATA 2222
Qy 869 TTACTTTTGTATCGTGCATCTGGTCAAGAAATTCCTATTGGGTGCACAAAGATTTTCCTG 928
Db 2223 ATAACATATTTTTTAAAGATTTTGTCTAAAAAAGAGATTTAGTTTGCACAAATTCATTG 2282
Qy 929 CTAGTCCAGATATCATGAGCCACAGTAAATTCGTTAAGTTTATATAATCCACACTTA 988
Db 2283 AGGATGCTTTTCTTAATGAACCAAGATAGATTTATTAAAGTATATAAGGATGAATTTG 2342
Qy 989 ACAACGCTGATAT 1001
Db 2343 GTAAATAGATTT 2355
RESULT 26
HSJ106901
```


LOCUS HSI106901 92357 bp DNA linear PRI 04-APR-2001
 DEFINITION Human DNA sequence from clone RP5-106901 on chromosome 20. Contains part of the gene for a novel protein similar to a low density lipoprotein-related protein LRP16 and STSs, complete sequence.
 ACCESSION AL049633.3 GI:4775651
 VERSION AL049633.3
 KEYWORDS HTG; LRP16.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 92357)
 Direct Submission
 Submitted (15-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 Requests: clonerquest@sanger.ac.uk
 On May 11, 1999 this sequence version replaced gi:4678553.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
 Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep/
 This sequence is the entire insert of clone RP5-106901 The true left end of clone RP4-738116 is at 92256 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr20>
 RP5-106901 is from the library RPCI-5 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
 VECTOR: PCYPAC2

FEATURES

source
 1. 92357
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="RZPD:RPCI7040011069"
 /db_xref="taxon:9606"
 /chromosome="20"
 /clone="RP5-106901"
 /clone_lib="RPCI-5"
 445..535
 /note="MIR repeat: matches 110..200 of consensus"
 repeat_region
 708..763
 /note="L2 repeat: matches 2653..2708 of consensus"
 repeat_region
 1438..1602
 /note="match: STS: Em:L16330"
 repeat_region
 1603..2171
 /note="MLR2CA repeat: matches 6..500 of consensus"
 repeat_region
 3883..4140
 /note="L2 repeat: matches 2480..2735 of consensus"
 repeat_region
 5015..5509
 /note="LTR37B repeat: matches 1..468 of consensus"
 repeat_region
 5707..5787
 /note="MIR repeat: matches 66..154 of consensus"
 repeat_region
 6118..6221
 /note="L1MC5 repeat: matches 7312..7409 of consensus"
 repeat_region
 6401..6736
 /note="L1MC5 repeat: matches 7560..7913 of consensus"
 repeat_region
 6803..6850

/note="24 copies 2 mer tt 75% conserved"
 6853..7163
 /note="AluSq repeat: matches 1..312 of consensus"
 repeat_region
 7333..7704
 /note="THE1C repeat: matches 1..371 of consensus"
 repeat_region
 7705..9306
 /note="THE1C-internal repeat: matches 1..1580 of consensus"
 repeat_region
 9307..9684
 /note="THE1C repeat: matches 1..371 of consensus"
 repeat_region
 9687..9770
 /note="5S repeat: matches 7..90 of consensus"
 repeat_region
 10529..10623
 /note="L2 repeat: matches 2642..2749 of consensus"
 repeat_region
 10724..10795
 /note="MIR repeat: matches 181..252 of consensus"
 misc_feature
 complement(11896..12355)
 /note="match: STS: Em:HSPP12C11 Em:294565"
 repeat_region
 12016..12114
 /note="MIR repeat: matches 36..130 of consensus"
 repeat_region
 12777..12823
 /note="MIR repeat: matches 77..123 of consensus"
 repeat_region
 12904..13054
 /note="MER5A repeat: matches 11..189 of consensus"
 repeat_region
 14132..14435
 /note="AluY repeat: matches 1..304 of consensus"
 repeat_region
 15118..15242
 /note="L2 repeat: matches 2125..2253 of consensus"
 repeat_region
 15257..15299
 /note="MER5B repeat: matches 7..49 of consensus"
 repeat_region
 15300..15351
 /note="13 copies 4 mer ttgtg 82% conserved"
 repeat_region
 15352..15460
 /note="MER5B repeat: matches 44..147 of consensus"
 repeat_region
 15460..15919
 /note="L2 repeat: matches 2275..2709 of consensus"
 repeat_region
 16557..16805
 /note="MIR repeat: matches 4..262 of consensus"
 repeat_region
 17067..17516
 /note="MLT2B repeat: matches 1..444 of consensus"
 repeat_region
 18905..18964
 /note="L2 repeat: matches 2689..2748 of consensus"
 repeat_region
 18914..18999
 /note="MIR repeat: matches 175..257 of consensus"
 repeat_region
 18990..19101
 /note="L2 repeat: matches 2598..2743 of consensus"
 repeat_region
 19107..19464
 /note="MLT1A1 repeat: matches 1..363 of consensus"
 repeat_region
 19482..19561
 /note="MER69 repeat: matches 2426..2505 of consensus"
 repeat_region
 19985..20378
 /note="L1MD2 repeat: matches 5686..6121 of consensus"
 repeat_region
 20395..20462
 /note="MER69 repeat: matches 1..68 of consensus"
 repeat_region
 21736..21783
 /note="12 copies 4 mer tcac 77% conserved"
 repeat_region
 21899..22178
 /note="AluSg1 repeat: matches 7..283 of consensus"
 repeat_region
 22507..22849
 /note="MER61B repeat: matches 7..425 of consensus"
 repeat_region
 23004..23082
 /note="L2 repeat: matches 2634..2709 of consensus"
 repeat_region
 23075..23372
 /note="L2 repeat: matches 2408..2697 of consensus"
 repeat_region
 23680..24013
 /note="L2 repeat: matches 2038..2384 of consensus"
 repeat_region
 25568..25664
 /note="LTR29 repeat: matches 515..617 of consensus"
 repeat_region
 25952..26055
 /note="LTR29 repeat: matches 26..221 of consensus"
 repeat_region
 26056..26150
 /note="MER34 repeat: matches 64..168 of consensus"
 repeat_region
 26540..28647

```
repeat_region /note="L1M43 repeat: matches 4167. .6304 of consensus"
29755. .30565
repeat_region /note="L1M44 repeat: matches 5473. .6300 of consensus"
31368. .31840
repeat_region /note="L1M37B repeat: matches 1. .468 of consensus"
31864. .33026
repeat_region /note="T1ggr3b repeat: matches 7. .1228 of consensus"
33105. .33174
repeat_region /note="L2 repeat: matches 2671. .2737 of consensus"
35600. .35623
repeat_region /note="12 copies 2 mer aa 100% conserved"
35695. .35860
repeat_region /note="M1R repeat: matches 23. .186 of consensus"
35966. .36119
repeat_region /note="M1R repeat: matches 36. .198 of consensus"
37193. .37331
repeat_region /note="MER58A repeat: matches 81. .224 of consensus"
37805. .37859
repeat_region /note="M1R repeat: matches 49. .115 of consensus"
37957. .38255
repeat_region /note="MLT1I repeat: matches 72. .410 of consensus"
39627. .39920
repeat_region /note="L2 repeat: matches 2362. .2707 of consensus"
41385. .41697
repeat_region /note="L1M16C repeat: matches 86. .387 of consensus"
41854. .42272
repeat_region /note="L2 repeat: matches 1377. .1805 of consensus"
42557. .42851
repeat_region /note="ALUSX repeat: matches 1. .295 of consensus"
43341. .43987
repeat_region /note="L1M16B repeat: matches 5488. .6171 of consensus"
44255. .44294
repeat_region /note="20 copies 2 mer ca 90% conserved"
44681. .45000
repeat_region /note="MER47A repeat: matches 29. .364 of consensus"
46252. .46605
repeat_region /note="MER63 repeat: matches 641. .988 of consensus"
46606. .46920
repeat_region /note="AluY repeat: matches 1. .304 of consensus"
46921. .46982
repeat_region /note="MER63 repeat: matches 988. .1050 of consensus"
47017. .47040
repeat_region /note="12 copies 2 mer tg 95% conserved"
47321. .47738
repeat_region /note="L1M16A repeat: matches 16. .448 of consensus"
47944. .48244
repeat_region /note="L1M4 repeat: matches 3552. .3871 of consensus"
48261. .48573
repeat_region /note="L1M16C4 repeat: matches 7218. .7522 of consensus"
49230. .49257
repeat_region /note="L2 repeat: matches 2626. .2653 of consensus"
49339. .49430
repeat_region /note="M1R repeat: matches 137. .227 of consensus"
49468. .49525

Query Match 4.3%; Score 47.4; DB 9; Length 92357;
Best Local Similarity 52.8%; Pred. No. 2.9;
Matches 102; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 736 GGAAGCGAAGTAGTATCCCGAGCGCTATAAAGTCACAGCAGATAGTCTACCTGCCACA 795
Db 6358 GAAAGCACAGACAGAAACCTACAGATTAAAGGACAGATATCTTGATTCAACAAA 6417
QY 796 TTATTAGAGTGGTTCACATGACCATTTGAAGCATCTCTGAGAAATCAGAAAGATAAT 855
Db 6418 CTATGTAAGAAAGTTACAAATTAAGAAATGAATAATCTATTATATCAAGGAAAGAAA 6477
QY 856 GTTAAATTTGATATTACTTTCTGTCGTGCTCTGTCAGAAATTCCTATTGGGTGCA 915
Db 6478 TTATATTAATTAATTTTCTAGTGTGTTAGTATTATTTTAACTATTCTATATTTTAGA 6537
QY 916 ACAAGATTTCTCTG 928
```

```
Db 6538 AATATATATACGG 6550

RESULT 27
AC008057 119082 bp DNA linear HTG 11-OCT-2000
LOCUS Homo sapiens chromosome 20 clone p1-10263, WORKING DRAFT SEQUENCE,
DEFINITION 3 unordered pieces.
AC008057
AC008057.4 GI:10765024
VERSION HTG; HTGS PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
On Oct 11, 2000 this sequence version replaced gi:9954568.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 68531: contig of 68531 bp in length
* 68532 102563: gap of unknown length
* 68532 102563: contig of 33932 bp in length
* 102564 102663: gap of unknown length
* 102664 119082: contig of 16419 bp in length.
FEATURES
Location/Qualifiers
source 1..119082
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosomes="20"
/clone="p1-10263"
BASE COUNT 37455 a 22465 c 21882 g 36187 t 1093 others
ORIGIN
Query Match 4.3%; Score 47.4; DB 2; Length 119082;
Best Local Similarity 52.8%; Pred. No. 2.9;
Matches 102; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 736 GGAAGCGAAGTAGTATCCCGAGCGCTATAAAGTCACAGCAGATAGTCTACCTGCCACA 795
Db 12709 GAAAGCACAGACAGAAACCTACAGATTAAAGGACAGATATCTTGATTCAACAAA 12768
QY 796 TTATTAGAGTGGTTCACATGACCATTTGAAGCATCTCTGAGAAATCAGAAAGATAAT 855
Db 12769 CTATGTAAGAAAGTTACAAATTAAGAAATGAATAATCTATTATATCAAGGAAAGAAA 12828
QY 856 GTTAAATTTGATATTACTTTCTGTCGTGCTCTGTCAGAAATTCCTATTGGGTGCA 915
Db 12829 TTATATTAATTAATTTTCTAGTGTGTTAGTATTATTTTAACTATTCTATATTTTAGA 12888
QY 916 ACAAGATTTCTCTG 928
Db 12889 AATATATATACGG 12901

RESULT 28
AC007630/c 169494 bp DNA linear PRI 22-NOV-2000
LOCUS AC007630
```

DEFINITION Homo sapiens chromosome 20 clone P1-7739, complete sequence.
ACCESSION AC007630
VERSION AC007630.4 GI:11245571
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 169494)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Direct Submission
JOURNAL Submitted (21-MAY-1999) Lita Annenberg Hazen Genome Sequencing Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA
REFERENCE 2 (bases 1 to 169494)
AUTHORS de la Bastide, M., Gnoj, L., Dedhia, N.N., Matero, A., Ning Huang, E., O'Shaughnessy, A., Preston, R., Rodriguez, M., Schutz, K., Shah, R., Shekher, M., Spiegel, L., Swaby, I., Vill, D. and McCombie, W.R.
TITLE Direct Submission
JOURNAL Submitted (02-JUL-1999) Lita Annenberg Hazen Genome Sequencing Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA
REFERENCE 3 (bases 1 to 169494)
AUTHORS McCombie, W.Richard.
TITLE Direct Submission
JOURNAL Submitted (18-NOV-2000) Cold Spring Harbor Laboratory
REMARK Tn10 Transposon removed.
REFERENCE 4 (bases 1 to 169494)
AUTHORS McCombie, W.Richard.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-2000) Cold Spring Harbor Laboratory
REMARK Tn10 Transposon removed.
COMMENT On Nov 21, 2000 this sequence version replaced gi:5332384. Clone P1-7739 contains a 3 kb overlap with Clone 48F5. The overlap ends at base 2940. From position 154930-155350 the subclones were single stranded and had single chemistry.

FEATURES
source
1..169494
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="20"
/clone="P1-7739"
misc_feature
134563
/note="Site of Tn10 transposon insertion. Tn10 Transposon, totalling 1328bp, was removed from final sequence."
/evidences=not experimental
BASE COUNT 52335 a 31812 c 32399 g 52939 t
ORIGIN

Query Match 4.3%; Score 47.4; DB 9; Length 169494;
Best Local Similarity 52.8%; Pred. No. 2.9;
Matches 102; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 736 GGAAGCGAAGATAGTATCCCGAGCGCTATAAAGTCACAGACATAGTCTACCTGCCACA 795
Db 12003 GAAGACACACAGAAACCTACAGATTAAAGGACAGATATCTTGATTCACAAACA 11944
QY 796 TTATTAGAAGTGGTTCACATGACCATTTGAAGCATTCCTGAAGAAATCAGAAAGATAAT 855
Db 11943 CTATGTAAGAAAGTTACATATTAATAATGAATAATACTTATATATCAAGGAAGAAA 11884
QY 856 GTTAAATGATATATCTTGTGTATCGTGATCTCGTCAAGAAATCTATTGGGTGCA 915
Db 11883 TTTAATATTAATTTTCAGTGTCTGTATAGTATTTATTTTAAACTATCTTATATTTTGA 11824
QY 916 ACAGATTTCCTG 928
Db 11823 AATATATATACGG 11811

RESULT 29

AE014817/c
LOCUS
DEFINITION Plasmodium falciparum 3D7 chromosome 14 section 2 of 13 of the complete sequence.
ACCESSION AE014817
VERSION AE014817.1
KEYWORDS GI:23497129
SOURCE Plasmodium falciparum 3D7
ORGANISM Plasmodium falciparum 3D7
REFERENCE 1 (bases 1 to 254449)
AUTHORS Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A., Nelson, K.E., Bowman, S., Paulsen, I.T., James, K., Eisen, J.A., Rutherford, K., Salzberg, S.L., Craig, A., Kyes, S., Chan, M.-S., Nene, V., Shallow, S.J., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M.W., Vaidya, A.B., Martin, D.M.A., Fairlamb, A.H., Fraunholz, M.J., Roos, D.S., Ralph, S.A., McFadden, G.I., Cummings, L.M., Subramanian, G.M., Mungall, C., Venter, J.C., Carucci, D.J., Hoffman, S.L., Newbold, C., Davis, R.W., Fraser, C.M. and Bartell, B.
TITLE Genome sequence of the human malaria parasite Plasmodium falciparum
JOURNAL Nature 419 (6906), 498-511 (2002)
PUBMED 12368864
REFERENCE 2 (bases 1 to 254449)
AUTHORS Gardner, M.J.
TITLE Direct Submission
JOURNAL Submitted (13-SEP-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

FEATURES
source
1..254449
/organism="Plasmodium falciparum 3D7"
/mol_type="genomic DNA"
/isoates="3D7"
/db_xref="taxon:36329"
/chromosome="14"
repeat_region
292..327
/rpt_type=tandem
repeat_region
712..780
/rpt_type=tandem
gene
<1089..>2856
/locus_tag="PF14_0066"
mRNA
join(<1089..>2095,2658..>2707,2795..>2856)
/locus_tag="PF14_0066"
CDS
join(1089..2095,2658..2707,2795..2856)
/locus_tag="PF14_0066"
/codon_start=1
/product="hypothetical protein, conserved",
/protein_id="AAN36678.1"
/db_xref="GI:23497130"
/translation="MEKSKRYISLIMMKERKFKYRLKQIMDNLKGIINKMKN
IPEIRRELKNIFFHNILSIKIKGLKYDRAYKVLFCQKNEKIATSLDFGSHKSLC
ISSQICGSGFCGKCATGQIGIKQLDIDITDOLLYFSQSGVDIKNISFMGGEPLAN
PVVFSIQPFNDNLFISNRRINISTVGLLFGIKLANNIPQVNLAFSLSHSPFTEER
DOLVINKLPFNEVFDLLDIRIAKTGRVIVISYILIKNLNDSKHAELASHICKRP
NNIRLYNVLCLIYKNGKRIYNIISPEYIVIVYLLIKKKILCKYIMFHTLYKYGIE
DINSFGYSIDAACGQIYADYEPKREKIESKMSLLL"
2114..2150
/rpt_type=tandem
repeat_region
2423..2460
/rpt_type=tandem
repeat_region
2918..2947
/rpt_type=tandem
repeat_region
2945..2973
/rpt_type=tandem
gene
<3572..>7390
/locus_tag="PF14_0067"
mRNA
<3572..>7390
/locus_tag="PF14_0067"
CDS
3572..7390
/locus_tag="PF14_0067"
/codon_start=1
/product="hypothetical protein"


```
Db 162536 --AAATGTTGATAATGTAATGTTGATAATGTAATGTTGATAATGTAATGTTGATAAT 162479
Qy 814 ATGACCATTAAGACATTCCTGAAGAATCAGAAAGATATGTTAAATTTGATTAAT 873
Db 162478 TAGATGTTGATAATGTAATGTTGATAATGTAATGTTGATAATGTAATGTTGATAAT 162419
Qy 874 TTGTTGATCGTATCGTCTGGTCAAGAAATTCATTGGGTGCAACAAGATTTCCTGCTAGT 933
Db 162418 GTAATGTTGATAATGTAATGTTGATAATGTAATGTTGATAATGTTGATAATGTTGATAAT 162359
Qy 934 CCAGAATATCAAGACCCACAGTTAAATTCCTTAAGTTTATAATCCACACTTTAAACAAC 993
Db 162358 GTAAATGTTGATAATGTAATGTTGATAATGTAATGTTGATAATGTAATGTAATGTAAT 162299
Qy 994 GGTGATATCCATATCAATATTAAGTTTTCAGCAACGCTTAGATGATGTCCTCCAGCT 1053
Db 162298 ATTAGTGTATAATGTTGATGCTCTTTATTAATATTTGGTTTCATGATGATGACGATGAT 162239
Qy 1054 CTCACCTGAAGGTATATAAAGAGGTAAACAAAATGTTAAGTATGTTG 1102
Db 162238 GATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 162190
```

```
RESULT 30
AC096684 84118 bp DNA linear HTG 20-NOV-2002
LOCUS Takifugu rubripes clone 241N7, WORKING DRAFT SEQUENCE, 2 ordered
DEFINITION pieces.
AC096684
VERSION AC096684.2 GI:25139901
KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT.
SOURCE Takifugu rubripes (Fugu rubripes)
ORGANISM Takifugu rubripes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae; Takifugu.
1 (bases 1 to 84118)
Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
Carliaga,K., Coleman,B., Engle,J., Granite,S., Guan,X., Gupta,J.,
Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E.,
Laric,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
Margulies,E.H., Masiello,C., Maskeri,B., McDowell,J.,
Pequirigan,C., Pearson,R., Portnoy,M.E., Prasad,A.,
Reddix-Dugue,N., Schandler,K., Schueler,M.G., Sison,C.,
Scantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 84118)
Green,E.D.
Direct Submission
Submitted (22-SEP-2001) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 84118)
Green,E.D.
Direct Submission
Submitted (20-NOV-2002) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
On Nov 20, 2002 this sequence version replaced gi:15721942.
```

```
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.zoo@nih.gov
----- Project Information
Center project name: arc
Center clone name: 241N07
```

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated

order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 83952 bases at least Q40
Consensus quality: 83954 bases at least Q30
Consensus quality: 83979 bases at least Q20
Insert size: 89000; agarose-fp
Insert size: 101000; pulse-field-gel
Insert size: 84018; sum-of-contigs
Quality coverage: 13.13x in Q20 bases; agarose-fp
Quality coverage: 11.57x in Q20 bases; pulse-field-gel
Quality coverage: 13.91x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 11009: contig of 11009 bp in length
* 11010 11109: gap of unknown length
* 11110 84118: contig of 73009 bp in length.

FEATURES

```
source
1..84118
/organism="Takifugu rubripes"
/mol_type="genomic DNA"
/db_xref="taxon:31033"
/clone="241N7"
/clone_lib="Incyte Genomics"
1..11009
/notes="assembly_fragment"
clone_end:17
vector_side:left
1..7443
/notes="clone overlaps with GenBank Accession Number
AC096997 clone 217A23 (center project name arb)"
11110..84118
/notes="assembly_fragment"
clone_end:SP6
vector_side:right

misc_feature
BASE COUNT 20759 a 20905 c 21005 g 21345 t 100 others
ORIGIN
```

```
Query Match 4.2%; Score 47; DB 2; Length 84118;
Best Local Similarity 44.6%; Pred. No. 3.6;
Matches 185; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

Qy 36 TGAAGTGCAAGCTGTTGTTAAACACATGCTCAGTTCAGATTCCAGATTAAAGGAGGTAC 95
Db 4887 TGTGTTGTTGCTGCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGCTGC 4946

Qy 96 AGCTTTGGTGAAGGTTGAGGCTGTTGCTGGTAACCAACTGATTGGAACATATTGCTTTA 155
Db 4947 TGTGTTGTTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGCT 5006

Qy 156 TAAAGTTGTCAGAAAGGTTCAATTCTAGATGTGACATGCTGTGACAGTTGTGTAACCT 215
Db 5007 TTTTGTGCTGCTGCTACTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGCTACTGT 5066
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2004, 05:08:13 ; Search time 376 Seconds
(without alignments)
7990.624 Million cell updates/sec

Title: US-10-081-644-1

Perfect score: 1113

Sequence: 1 atgcagttcccaaccatca.....agtatgtgccaggtataa 1113

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : N Geneseq_19Jun03.*

```

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	357.8	32.1	1631	24	ABQ76453
2	98.4	8.8	393	24	ABN26679
3	71	6.4	1086	24	ABQ76566
4	64.4	5.8	1330	24	ABT11258
5	45.4	4.1	23439	18	AAV74349
6	40.8	3.7	3501	21	AAA70202
7	40.2	3.6	439	22	ABA49878
8	40.2	3.6	439	22	ABA67796

Probe #13320 for 9	40.2	3.6	439	22	ABA34854
Human brain expres	40.2	3.6	439	22	AAK16207
Human bone marrow	40.2	3.6	439	22	AAK41953
Probe #12651 for 9	40.2	3.6	439	22	AAI22718
Probe #16702 used	40.2	3.6	439	22	AAI48016
Probe #8374 used c	40.2	3.6	439	22	AAI08383
Human liver single	40.2	3.6	439	23	ABS41558
Human genome-deriv	40.2	3.6	439	24	ABS15980
Human huntington (40	3.6	1236	24	AAI44410
AmEPV first RNA po	39.6	3.6	1050	24	ABL56207
Human immune/haema	39.6	3.6	4027	22	AAK67473
AmEPV genome fragm	39.6	3.6	50000	24	ABL55644
Strawberry fruit a	39.4	3.5	772	18	AAI91162
Streptococcus poly	39.4	3.5	1035	24	ABN66435
Strawberry fruit a	39.4	3.5	1151	18	AAI91136
Streptococcus poly	39.4	3.5	215561	24	ABN71527
Human GDP-mannose	39.4	3.5	392	25	ABX19561
Genomic sequence o	39.4	3.5	236589	24	ABA90521
Gastric cancer ass	38.8	3.5	847	20	AAK39814
Polyglutamine trac	38.4	3.4	486	22	AAF75507
Bacteriophage 96 c	38.2	3.4	43576	21	AAA68609
AmEPV genome fragm	38	3.4	50000	24	ABL55643
S. pneumoniae type	37.6	3.4	987	25	ABX06427
L-lactic acid dehy	37.6	3.4	3073	12	AAQ15280
Streptococcus pneu	37.6	3.4	640681	24	ABA92787
Buchnera sp. genom	37.6	3.4	2162598	25	ABS56454
Streptococcus pneu	37.4	3.4	4920	23	ABL08746
Drosophila melanog	37.4	3.4	23761	24	ABQ80552
Human Canion gene	37.2	3.3	860	22	AAH03300
Human cDNA clone (37.2	3.3	1350	23	AAI72599
DNA encoding novel	37.2	3.3	2792	21	AAA26994
Maize PCNA P120 ge	37.2	3.3	1163020	24	ABQ67197
Listeria innocua c	37.2	3.3	3011208	24	ABQ69245
Listeria innocua D	37	3.3	1755	24	ABN68175
Streptococcus poly	37	3.3	1923	23	ABL28551
Drosophila melanog	37	3.3	3582	21	AAI70341
Plasmodium falci	37	3.3	7008	22	AAI46531
Tumour suppressor	37	3.3	8772	22	AAI11678
Human Flamingo CDN	37	3.3	8871	22	AAI11677
Human Flamingo CDN	37	3.3	9321	21	AAI76401
Human ORFX ORF1956	37	3.3	9321	21	AAI76401
Human REPTX 9 CDNA	37	3.3	9321	21	AAI76401
Human GPCR CELSR2	37	3.3	10531	25	ABZ42868
Human mdt cDNA SE	37	3.3	11762	25	ABX34546
Human cDNA differe	37	3.3	198285	24	ABK84699
Gene #3817 used to	37	3.3	198285	24	ABN97319
Borrelia burgdorfe	37	3.3	910715	20	AAI20248
Streptococcus poly	37	3.3	215561	24	ABN71527
Human breast cell	36.8	3.3	441	22	ABA44736
Human foetal liver	36.8	3.3	441	22	ABA55191
Probe #1403 for ge	36.8	3.3	441	22	ABA24937
Human brain expres	36.8	3.3	441	22	AAK03451
Human bone marrow	36.8	3.3	441	22	AAK28907
Probe #3426 for ge	36.8	3.3	441	22	AAI13493
Probe #3537 used t	36.8	3.3	441	22	AAI34851
Probe #3366 used t	36.8	3.3	441	22	AAI03375
Human liver single	36.8	3.3	441	23	ABS28520
Human genome-deriv	36.8	3.3	441	23	ABS03436
Arabidopsis thalia	36.8	3.3	3150	24	ABL21438
Drosophila melanog	36.8	3.3	4629	23	ABL29279
Drosophila melanog	36.8	3.3	6693	23	ABL29278
Enterococcus faeca	36.6	3.3	969	23	AAI52969
Drosophila melanog	36.6	3.3	2240	23	ABL28548
Drosophila melanog	36.6	3.3	4324	23	ABL28550
Enterococcus faeca	36.6	3.3	6253	20	AAI13097
Enterococcus faeca	36.6	3.3	6253	24	ABS98892
Methanococcus jann	36.6	3.3	1664976	19	AAV21209
Sequence encoding	36.6	3.3	4590	7	AAI60472
Bacteriophage 3A c	36.4	3.3	43095	21	AAA68254
Arabidopsis thalia	36.2	3.3	1323	21	AAI32804
Nucleic acid seque	36.2	3.3	3255	20	AAI99608
Mycoplasma genital	36.2	3.3	580073	18	AAI58840

c	82	36	3-2	222	21	ABN81097	Shrimp polynucleot
	83	36	3-2	312	23	AA533374	Haemophilus influ
	84	36	3-2	978	22	AAH50794	CFE 98 coding seq
c	85	36	3-2	9095	24	ABK28448	DNA transcription
c	86	36	3-2	40324	24	ABQ67150	Human angiogenesis
c	87	36	3-2	1082138	21	AAAT23305	Arabidopsis thaliana
	88	36	3-2	1830121	17	AAAT42063	Haemophilus influe
	89	35.8	3-2	840	20	AAZ211567	Staphylococcus aur
	90	35.8	3-2	840	20	AAZ211567	S. aureus spo02 p
	91	35.8	3-2	840	20	AAZ29397	

ALIGNMENTS

RESULT 1

RESUL I
AB076453

ABQ76453
ID ABQ76453 standard; cDNA: 1631 BP.

AC ABQ76453;

DT 21-NOV-2002 (first entry)

DE S. cerevisiae BAX-associated cDNA fragment SEQ ID 331.

Bax; Bax-resistance; cytostatic; fungicide; immunosuppressive; virucide;
 KW vasotropic; vaccine; gene therapy; proliferative disorder; cancer;
 KW apoptosis; fungal; yeast; infection; autoimmune disease; ischaemia;
 KW neurodegeneration; cell death; ss.

OS *Saccharomyces cerevisiae*.

XX
PN
WO200264766-A2.

22-AUG-2002.

21-DEC-2001; 2001WO-EP15398.

AA
PR 22-DEC-2000; 2000EP-0870318.

PR 04-JAN-2001; 2001EP-08700002.

PR 09-JAN-2001; 2001EP-0870003.

PA (JANC) JANSSEN PHARM NV.

PI Contreras RH, Eberhardt I, Luyten WHML, Reekmans RJ;

WPI; 2002-667002/71.

DR P-PSDB; ABG93187.

PT New isolated nucleic acid representing a synthetic Bax gene, useful as
PT medicament for treating, preventing and/or alleviating yeast or fungal
PT infections or proliferative disorders, or for preventing apoptosis in
PT certain diseases -

PS Claim 36; Figure 1; 344pp; English.

This invention describes a novel nucleic acid representing a synthetic Bax gene. The Bax gene of the invention is useful for identifying Bax-resistant yeast or fungi, identifying, or obtaining and identifying *Candida* spp. sequences that are differentially expressed in a pathway eventually leading to programmed cell death or identifying inhibitors or inhibitor sequences of Bax-induced cell death. The products of the invention have cytostatic, fungicide, immunosuppressive, virucide and vasotropic activity and can be used in vaccines or for gene therapy. The isolated nucleic acids, polypeptides, pharmaceutical compositions, antisense molecules and antibodies are useful as medicaments or in preparing a medicament for treating, preventing and/or alleviating diseases associated with yeast or fungi or proliferative disorders, such as cancer, or for preventing apoptosis in certain diseases. The compound or polypeptides, or the genetically modified organism are useful for preparing a medicament for modifying the endogenous flora of humans or other mammals. The vaccine is useful for immunising against yeast or fungal infections. Apoptosis-related diseases include autoimmune diseases, ischaemia, diseases related with viral infections or neurodegenerations.

PI Contreras RH, Eberhardt I, Luyten WHML, Reekmans RJ;
XX WPI: 2002-667002/71.
DR P-PSDB; ABG93300.
XX
XX New isolated nucleic acid representing a synthetic Bax gene, useful as
PT medicament for treating, preventing and/or alleviating yeast or fungal
PT infections or proliferative disorders, or for preventing apoptosis in
PT certain diseases -
XX
XX Claim 36; Figure 2; 344pp; English.
XX
XX This invention describes a novel nucleic acid representing a synthetic
CC Bax gene. The Bax gene of the invention is useful for identifying
CC Bax-resistant yeast or fungi, identifying, or obtaining and identifying
CC Candida spp. sequences that are differentially expressed in a pathway
CC eventually leading to programmed cell death or identifying inhibitors or
CC inhibitor sequences of Bax-induced cell death. The products of the
CC invention have cytostatic, fungicide, immunosuppressive, virucide and
CC vasotropic activity and can be used in vaccines or for gene therapy. The
CC isolated nucleic acids, polypeptides, pharmaceutical compositions,
CC antisense molecules and antibodies are useful as medicaments or in
CC preparing a medicament for treating, preventing and/or alleviating
CC diseases associated with yeast or fungi or proliferative disorders, such
CC as cancer, or for preventing apoptosis in certain diseases. The compounds
CC or polypeptides, or the genetically modified organism are useful for
CC preparing a medicament for modifying the endogenous flora of humans and
CC other mammals. The vaccine is useful for immunising against yeast or
CC fungal infections. Apoptosis-related diseases include autoimmune disease,
CC ischaemia, diseases related with viral infections or neurodegenerations.
CC This sequence represents a polynucleotide associated with the Bax gene
CC described in the disclosure of the invention.
XX
XX Sequence 1086 BP; 359 A; 174 C; 223 G; 330 T; 0 other;
SQ

Query Match 6.4%; Score 71; DB 24; Length 1086;
Best Local Similarity 49.0%; Pred. No. 4.2e-09;
Matches 324; Conservative 0; Mismatches 310; Indels 27; Gaps 4;
35 TTGAGGTGCAAGCTGTTGTTAAACAGAGTGTCTCAGTTCAGAAATTAAGAGGGTA 94
Db 29 TCGAACCTTATCATATTAGCGGAAATTAAGATATTCCTCAACAAAATAAAGAAATG 88
95 CAGCTTTGGTGAAGTTGAGCTGTCTGTGTACCACTGATTGGAAGCATATTCCTT 154
Db 89 AATATTATCAAGCAGTAGCTTTTGCATTAACCACTGATTGGAAGCATATTCCTT 148
95 ATAAGATTGTTCCAGAAAGTTCAATTTAGGAGTGACATTTGCTGGTACAGTTGTCAAAC 214
Db 149 ATCAATTTGGGCAGCCAGGTGATGTTGTTGCGGATGTTAGTGGGATCATTTGAAGAAG 208
95 TTGGACCAATGTAGTACTGTGACTTTGAAGTTGAGATACGGTTTCGGTTTGTTCACG 274
Db 209 TGGGTTCTCAAG-----TAACTGGGTTTGCAAAAGGTGACACTGTAAGTGTCTTTATAC 265
95 GTGCTTCCCAACAGATCTTAAATGCGTCAATTTGCTGAATATGCCAGGTTTATCCAC 334
Db 266 GTA-----ATAGATCACTCGCACTGGAGCTTTTGAGAAATATGTAGCTGTGATCTG 319
95 CTTTGTGTTTCAAGAGTAACCTTAACCTCACTCACTGCTGATGAATTTCTGAAGGCCCTG 394
Db 320 CTACTTCGATAAAGTACAATAAGAATTTGAACATTTTGACTAATTTACAAGTATCTGAAA 379
95 TGAGAACTTCGATCTGCTGCATCATTTGCCAGTTTCGTTGACAACTGCTGGTTAGTT 454
Db 380 TCCACTCATTTGAAGGGGCGAGCAAGTATTAATTTAGGTTTGGTTACCGTTGGGCTTCAT 439
95 TGTGTCACTACTTGGGCTCAAAATATGAATGGCACTTACCCGCAACATCTATCATC 514
Db 440 TTCTCTACTTACGAATTTGACAAACAAAGCAACC-----TGGGGATA 484
95 CATTTATTTGTTGGGTTGGTGTCTACAGCAGTGGGTCAACAACTAATCAAGTTGCCAAAC 574

485 GTATTTTGTATTTGGGGAGGAGCAACTGCAACTGGAGTTCTAGCCATTCAGTTGCCAAAC 544
Qy 575 ATATCAATGCTTATCTAAGATTGTAACTGCTCTTCTTAAAGAGCATGAAAGCTTTTAA 634
Db 545 TAGTGTA---TAATCTCAAGTAATCACCACAGCATCACCCAAAACCCACCCCTTTGA 601
Qy 635 AGTCTTTATGCTGTGATGATGCTTTTGACTATCATGATGAGCGGTTATTTGACGATCA 694
Db 602 AACAAATTAGGGCGAGATTAATGTTTTCGATTATGAGACGCTGATTTGTCAATAAATA 661
Qy 695 A 695
Db 662 A 662

RESULT 4
ABT11258
ID ABT11258 standard; DNA; 1330 BP.
XX
XX AC ABT11258;
XX
XX DT 10-DEC-2002 (first entry)
XX
XX Yeast selected interacting domain coding sequence SEQ ID NO: 15.
XX
XX Yeast; protein-protein interaction; Selected Interacting Domain;
KW SID (RTM); secretion yield; cancer; neurodegenerative disease; fungicide;
KW cytostatic; neuroprotective; gene; ds.
XX
XX Saccharomyces cerevisiae.
XX
XX WO200266504-A2.
XX
XX PD 29-AUG-2002.
XX
XX 14-FEB-2002; 2002WO-EP02299.
XX
XX PR 16-FEB-2001; 2001US-269266P.
XX
XX (HYBR-) HYBRIGENICS.
XX
XX PI Legrain P;
XX
XX WPI: 2002-674913/72.
DR P-PSDB; ABJ10941.
XX
XX New protein-protein complexes of Saccharomyces cerevisiae, useful in
PT drug screening or development, for developing yeast strains with better
PT secretion yield of protein, or in gene therapy (e.g. to treat Candida
PT infection or cancer) -
XX
XX Claim 7; Page 41; 357pp; English.
XX
XX The present invention relates to complexes between Saccharomyces
CC cerevisiae Selected Interacting Domain (SID (RTM)) proteins and coding
CC sequences. The protein complexes of S. cerevisiae are useful in drug
CC development, in screening drugs or agents that modulate the interaction
CC of proteins, in developing yeast strains with better secretion yield of
CC protein, and in gene therapy. The protein complexes, polypeptides and
CC polynucleotides are useful for preventing or treating Candida infection,
CC cancer or neurodegenerative diseases in humans or animals. The present
CC sequence is a coding sequence of the invention.
XX
XX Sequence 1330 BP; 420 A; 257 C; 207 G; 446 T; 0 other;
SQ

Query Match 5.8%; Score 64.4; DB 24; Length 1330;
Best Local Similarity 56.0%; Pred. No. 2.9e-07;
Matches 122; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
895 CAAGAATTTCTATTGGGTGCAACAGATTTCTGCTAGTCCAGAAATATCATGAGCCACA 954
Qy 1 CATGAAGTACCATTGGAACATTTACTTTACCAGCGCATTTACTTACCAGCGCATTTAGGAAGCTGCA 60
Db

Qy	955	GTTHAAATTCGGTTAAAGTTTATAAAATCCACACCTTAAACACGGTGATATGCCATCATATGAAT	1014
Db	61	ATRAAAATTTATCAAATTCATCAATCCAAAGATTAAATGATGGACAAATTCGCCATATTCCCA	120
Qy	1015	ATTTAAAGTTTTTCAGCAACGGCTTAGATGATGTCCAGCTCTCACTGAGGTATATAAAGAA	1074
Db	121	GTAAGGGTCATATAAGAACGGGCTTTGTGATGTTCCCTCATATCCCTAAAGACATCAAAATAT	180
Qy	1075	GGTAAAAACAAAAATGTTAAAGTATGTTGCCAGGTTTATA	1112
Db	181	GGTAAGAAGTCTCGTGAAAAACTCGTTGCCGTATTAAA	218
RESULT 5			
AAV74349/c			
XX	AAV74349 standard; DNA; 23439 BP.		
XX	AAV74349;		
XX	16-MAR-1999 (first entry)		
XX	Staphylococcus aureus contig SEQ ID #38.		
XX	Computer readable medium; vaccine; S.aureus infection; immunodetection;		
KW	cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;		
KW	skin infection; surgical wound infection; scalded skin syndrome;		
XX	toxic shock syndrome; ds.		
XX	Staphylococcus aureus.		
XX	Key		
FH	Location/Qualifiers		
FT	misc_feature	481..540	
FT	/*tag= a		
FT	/note= "these bases represent a line of missing text in		
FT	the sequence listing in the specification. They		
FT	are included to maintain the nucleotide numbering		
FT	given in the specification for this DNA sequence"		
FT	misc_feature	2281..2340	
FT	/*tag= b		
FT	/note= "these bases represent a line of missing text in		
FT	the sequence listing in the specification. They		
FT	are included to maintain the nucleotide numbering		
FT	given in the specification for this DNA sequence"		
FT	misc_feature	4081..4140	
FT	/*tag= c		
FT	/note= "these bases represent a line of missing text in		
FT	the sequence listing in the specification. They		
FT	are included to maintain the nucleotide numbering		
FT	given in the specification for this DNA sequence"		
FT	misc_feature	5881..5940	
FT	/*tag= d		
FT	/note= "these bases represent a line of missing text in		
FT	the sequence listing in the specification. They		
FT	are included to maintain the nucleotide numbering		
FT	given in the specification for this DNA sequence"		
FT	misc_feature	7681..7740	
FT	/*tag= e		
FT	/note= "these bases represent a line of missing text in		
FT	the sequence listing in the specification. They		
FT	are included to maintain the nucleotide numbering		
FT	given in the specification for this DNA sequence"		
FT	misc_feature	9481..9540	
FT	/*tag= f		
FT	/note= "these bases represent a line of missing text in		
FT	the sequence listing in the specification. They		
FT	are included to maintain the nucleotide numbering		
FT	given in the specification for this DNA sequence"		
FT	misc_feature	11281..11340	
FT	/*tag= g		
FT	/note= "these bases represent a line of missing text in		
FT	the sequence listing in the specification. They		
FT	are included to maintain the nucleotide numbering		
FT	given in the specification for this DNA sequence"		

CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the *S.aureus* DNA sequences contained on the
CC computer readable medium.

Seq	Sequence	23439 BP;	7768 A;	3506 C;	4455 G;	6921 T;	789 other;
	Query Match	4.1%;	Score 45.4;	DB 18;	Length 23439;		
	Best Local Similarity	45.4%;	Pred. No. 0.14;				
	Matches 163;	Conservative 0;	Mismatches 196;	Indels 0;	Gaps 0;		
Qy	504	ACATACTCATCCATTATTGATTTGGGGTGTGTCACAGCATGGGTCAACAACTAATCCA	563				
Db	16078	ACAACGTTTTTAAATCTACGTTTTATGCACGTGCCACAGTAATAATGGATAAATATGATTTAA	16019				
Qy	564	AGTTGCCAAACATATCAATGCTTTATACTAAGATTGTAACTGTGTCTTTAAAAAGCATGA	623				
Db	16018	CATTGCTGATTACTTAATTCAAAATATATATTGCCACAGATGACTTTAAACACTTCC	15959				
Qy	624	AAAGCTTTTAAAGCTTTATGGTGTCTGATGATGCTTTTGACTATCATGATCGAGGGTTAT	683				
Db	15958	ACCTGTTTTTCATTGGCGATTGTAATCGCATATATGATGTTCCGTTTGAGAAAGCGAACA	15899				
Qy	684	TGACGAGATCAAAATCGAAGTATCCAAACCTCGCAACATGTTATTGACGCTGTGGGAAGCGA	743				
Db	15898	TATCATGAATCATGTTCCACATTTCCACATTTGAAACGCGTGAATAAAAACGACATGATTT	15839				
Qy	744	AGATAGTATCCCCGAGGCCCTATAAAGTCTACAGCAGATAGTCTACCTGCCACATTAATTAGA	803				
Db	15838	TGATCGTAGACCTTAATGACGAAGCAATCACTATTTATCGTAAAGTTGTTGATTTCTTAAA	15779				
Qy	804	AGTGGTTCCATGACCATTTGAAGCAATTCCTGAGAAATCGAAGAAATATGTTAAAAA	862				
Db	15778	CGCTATAACCAATGATGTAAATATATCATTAATTTTAAAAAACTCAATTTATCATCATGATA	15720				

RESULT 6	
AAA70202	AAA70202 standard; DNA; 3501 BP.
ID	AAA70202 standard; DNA; 3501 BP.
XX	
XX	AAA70202;
AC	
DT	07-NOV-2000 (first entry)
XX	
DE	Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:335.
XX	
XX	Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW	antimalarial; malaria; protozoasidae; infection; insecticide; ds.
XX	
OS	Plasmodium falciparum.

The present invention describes proteins and their fragments (I) encoded by chromosome 2 of the human malarial parasite, *Plasmodium falciparum*. Also described are: (I) nucleotide sequences (II) encoding (I); and (2) vaccines against *P. falciparum* infection comprising (I) or (II). (I) and (II) are useful for the development of vaccines against *P. falciparum* infection. (I) and polyclonal antisera or a monoclonal antibody raised to immunogens comprising the sequences of (I), are useful in the detection of infection with *P. falciparum*. Furthermore, (I) (especially when they are rifins or secreted or membrane proteins) can aid the identification of drugs to treat or prevent *P. falciparum* infection, or they can be used to identify drug resistance in *P. falciparum*. Sequencing of the *Plasmodium* chromosome 2 and the subsequent identification of proteins encoded by it will help to expand our understanding of parasite biology, a process hampered by the complexity of the parasitic lifecycle, and provide new targets for vaccine and drug development. Parasite resistance to drugs and mosquito resistance to insecticides have led to a resurgence of malaria in many parts of the world, and there is a pressing need for vaccines and new drugs. AAA70078 to AAA70287 and AAA1814 to AAA18352 represent nucleotide and protein sequences given in the present invention, but which are not specifically mentioned within the specification.

	SQ	Sequence	3501 BP; 1699 A; 327 C; 468 G; 1007 T; 0 other;	
		Query Match	3.7%; Score 40.8; DB 21; Length 3501;	
		Best Local Similarity	46.2%; Pred. No. 1-2;	
		Matches	209; Conservative 0; Mismatches 237; Indels 6; Gaps 2;	
	Qy	650	ATGATGTCTTTGACTATCATGATGCAGGGCGTATTTCAGCAGATCAAAATCGAAGTATCCAA	709
	Db	194	ATGATGAATCTGTTTAAGACGCACGTCAAAGTGAAATGAGGAGATCCAACAAGAAGAAAACGA	253
	Qy	710	ACCTGCCAACATGTTATTTCAGCTGTGGGAGCGAAGATAGTATCCCCGAGGCCCTATAAG	769
	Db	254	ACCATGATATATAAAACAATAATGATGTAATGATGATAATAT---AAATAATAAATCTCTA	310
	Qy	770	TCACAGCAGATGCTTACCTGCCACATTTATTAGAAGTGGTTCCTCAATGACCATTGAAAGCA	829
	Db	311	TAAATGATGATAATATAAGCAATAATCATATAGTAGTATGATATATAAACAATAATCATATA	370
	Qy	830	TTCCCTCAAGAAAAATCAGAAAAGATAATGTTTAAAAATTGATATTACTTTGTGTATCGTGCAT	889
	Db	371	TAAATGATGATAATATAAACAATAATGATGTAATGATGATATAATAATAATAATCAATT	430
	Qy	890	CTGGTCAGAAATCTTATTTGGGGTGCAACAAGATTTCCCTCGTAGTCAGAGATATCATGAAG	949
	Db	431	TAAATGATGATAATATAAAATAATAATCATATAAATGATGATATATAAAAATAATAACCATA	490
	Qy	950	CCACAGTTAAATTCGTTTTAGTTTTATAAAATCCACACCTTAACACGGTGATATCCATCATA	1009
	Db	491	TAAATGATGATAATATAAACAATAATAACCATATAAATGATGATATAATAATAATAATTATT	550
	Qy	1010	TGAATATTAAGTTTTTCAGCACCGGCTTAGATG---ATGTCACCGCTCTCACTGGAAGGTA	1066
	Db	551	GTAATATGATGATATGATATACATATCAANTGTTCCATATCATCTGTAATAATTCTCCCTA	610
	Qy	1067	TAAAGAAGGTAAAAACAATAATGTTAAGTAT	1098
	Db	611	AAAAAGAAAAGAAAAGAAAATGTTAGANTAT	642

RESULT 7	
ABA49878/c	
ID ABA49878	standard; DNA; 439 BP.
XX	
XX	
ABA49878;	
XX	
XX	
DT 01-FEB-2002	(first entry)
XX	
DE	Human breast cell single exon nucleic acid probe #8573.
XX	
KW	Human; microarray; single exon probe; gene expression; breast;
KW	disease; cancer; ss.
KW	

XX OS Homo sapiens.
XX PN WO200157271-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00662.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX PD WPI; 2001-496933/54.
XX PT New spatially-addressable set of single exon nucleic acid probes,
XX PR useful for measuring gene expression in sample derived from human
XX PR breast, comprises number of single exon nucleic acid probes -
XX PS Claim 4; SEQ ID NO 8573; 327pp + sequence listing; English.
XX CC The invention relates to a spatially-addressable set of single exon
XX CC nucleic acid probes for measuring gene expression in a sample derived
XX CC from human breast and BT 474 cells. The method involves contacting
XX CC the probes with a collection of detectably labelled nucleic acids
XX CC derived from mRNA of human breast, and then measuring the label
XX CC bound to each probe of the microarray. The probes are useful for
XX CC verifying the expression of regions of genomic DNA predicted to
XX CC encode proteins. They are useful for gene discovery, and for
XX CC determining predisposition and/or prognosing breast disease. Gene
XX CC expression analysis is useful for assessing the toxicity of chemical
XX CC agents on cells. The microarray of this invention presents a far greater
XX CC diversity of probes for measuring gene expression, with far less bias
XX CC than expressed sequence tag microarrays. The method is suitable for
XX CC rapid production of functional information from genomic sequence. The
XX CC present sequence is a single exon nucleic acid probe of the invention.
XX CC Note: The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 439 BP; 141 A; 169 C; 41 G; 88 T; 0 other;

Query Match 3.6%; Score 40.2; DB 22; Length 439;
Best Local Similarity 46.8%; Pred. No. 0.74; Indels 0; Gaps 0;
Matches 126; Conservative 0; Mismatches 143; Indels 0; Gaps 0;
QY 36 TGAAGGTGACAAAGCTGTTGTTAAACAGATGCTCAGTTCCAGATTAAGAGGGGTAC 95
DB 385 TGATGGTGGTGATGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATG 326
QY 96 AGCTTGGTGAAGTTGAGGCTGTTGCTGCTAACCCAACTGATTGGAAGCATATTGCTTA 155
DB 325 TGCCGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 266
QY 156 TAAGATTGGTCCAGAAAGTTCAATTCAGGATGTGACATTTGCTGGTACAGTTGTCAAACT 215
DB 265 TGCTGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 206
QY 216 TGGACCAAAATGCTAGTACTGACTTTGAAGGTTGGAGATACCGGTTTTCGTTTTCACGG 275
DB 205 TCCTGCTGATGGTTTTCATGTTATGATGATGATGATGATGATGATGATGATGATGATG 146
QY 276 TGCTTCCCAACAGATCCTAAAATGGTG 304
DB 145 TGATGATGATGGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 117

RESULT 8
ABA67796/c
ID ABA67796 standard; DNA; 439 BP.
XX AC ABA67796;
XX DT 01-FEB-2002 (first entry)
XX DE Human foetal liver single exon nucleic acid probe #16101.
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; sb.
XX OS Homo sapiens.
XX PN WO200157277-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00669.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX PD WPI; 2001-483447/52.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PR analyzing gene expression in human fetal liver -
XX PS Claim 4; SEQ ID NO 16101; 639pp + sequence listing; English.
XX CC The invention relates to a single exon nucleic acid probe for
XX CC measuring human gene expression in a sample derived from human foetal
XX CC liver. The single exon nucleic acid probes may be used for predicting,
XX CC measuring and displaying gene expression in samples derived from human
XX CC foetal liver. The present sequence is a single exon nucleic acid
XX CC probe of the invention.
XX CC Note: The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 439 BP; 141 A; 169 C; 41 G; 88 T; 0 other;

Query Match 3.6%; Score 40.2; DB 22; Length 439;
Best Local Similarity 46.8%; Pred. No. 0.74; Indels 0; Gaps 0;
Matches 126; Conservative 0; Mismatches 143; Indels 0; Gaps 0;
QY 36 TGAAGGTGACAAAGCTGTTGTTAAACAGATGCTCAGTTCCAGATTAAGAGGGGTAC 95
DB 385 TGATGGTGGTGATGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATG 326
QY 96 AGCTTGGTGAAGTTGAGGCTGTTGCTGCTAACCCAACTGATTGGAAGCATATTGCTTA 155
DB 325 TGCCGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 266
QY 156 TAAGATTGGTCCAGAAAGTTCAATTCAGGATGTGACATTTGCTGGTACAGTTGTCAAACT 215
DB 265 TGCTGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 206
QY 216 TGGACCAAAATGCTAGTACTGACTTTGAAGGTTGGAGATACCGGTTTTCGTTTTCACGG 275
DB 205 TCCTGCTGATGGTTTTCATGTTATGATGATGATGATGATGATGATGATGATGATGATG 146
QY 276 TGCTTCCCAACAGATCCTAAAATGGTG 304
DB 145 TGATGATGATGGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 117

CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 439 BP; 141 A; 169 C; 41 G; 88 T; 0 other;

Query Match 3.6%; Score 40.2; DB 22; Length 439;
Best Local Similarity 46.8%; Pred. No. 0.74;
Matches 126; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

QY 36 TGAAGGTGACAAAGCTGTTGTTAAACAGATGCTCAGTTCACAGATTAAGGAGGCTAC 95

DB 385 TGATGGTGGTATGCTGCTGATGATGCTGCTGATGATGCTGCTGATGATGCTGCTGATG 326

QY 96 AGCCTTGGTGAAGGTTGAGGCTGTTGCTGTTAAACAGATGCTCAGTTCACAGATTAAGGAGGCTAC 155

DB 325 TGCCGATGGTATGCTGCTGATGATGCTGCTGATGATGCTGCTGATGATGCTGCTGATG 266

QY 156 TAAGATTGGTCCAGAGGTTCAATCTAGGATGATGATGATGATGATGATGATGATGATGATG 215

DB 265 TGCTGATGGTATGCTGCTGATGATGCTGCTGATGATGCTGCTGATGATGCTGCTGATG 206

QY 216 TGGACCAAACTAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 275

DB 205 TCCTGCTGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 146

QY 276 TGCTTCCCAACACAGATCCTAAAAATGGTG 304

DB 145 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 146

RESULT 15

ABS41558/c

ID ABS41558 standard; DNA; 439 BP.

XX AC ABS41558;

XX 25-FEB-2003 (first entry)

XX Human liver single exon probe, SEQ ID No 16548.

XX Human; single exon nucleic acid probe; liver; cirrhosis;

KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;

KW coronary heart disease; ss.

XX Homo sapiens.

XX WO200157273-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00664.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024253.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488898/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analysing gene expression in human adult liver -

XX Claim 4; SEQ ID No 16548; 658pp; English.

XX The invention relates to a single exon nucleic acid probe (SENP) (I) for

CC measuring human gene expression in a sample derived from human adult

CC liver, comprising one of 13109 defined nucleotide sequences given in the

CC specification (or complements/fragments). The probe hybridizes at high

CC stringency to a nucleic acid molecule expressed in the human adult

CC liver. (I) may be used for predicting, measuring and displaying gene

CC expression in samples derived from human adult liver. The genes

CC identified may be involved in genetic liver diseases such as cirrhosis,

CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which

CC is associated with coronary heart disease. ABS25011-ABS51005 represent

CC human liver single exon nucleic acid probes of the invention.

CC Note: The sequence information for this patent does not appear in the

CC printed specification but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 439 BP; 141 A; 169 C; 41 G; 88 T; 0 other;

Query Match 3.6%; Score 40.2; DB 23; Length 439;

Best Local Similarity 46.8%; Pred. No. 0.74; Indels 0; Gaps 0;

Matches 126; Conservative 0; Mismatches 143;

QY 36 TGAAGGTGACAAAGCTGTTGTTAAACAGATGCTCAGTTCACAGATTAAGGAGGCTAC 95

DB 385 TGATGGTGGTATGCTGCTGATGATGCTGCTGATGATGCTGCTGATGATGCTGCTGATG 326

QY 96 AGCCTTGGTGAAGGTTGAGGCTGTTGCTGTTAAACAGATGCTCAGTTCACAGATTAAGGAGGCTAC 155

DB 325 TGCCGATGGTATGCTGCTGATGATGCTGCTGATGATGCTGCTGATGATGCTGCTGATG 266

QY 156 TAAGATTGGTCCAGAGGTTCAATCTAGGATGATGATGATGATGATGATGATGATGATGATG 215

DB 265 TGCTGATGGTATGCTGCTGATGATGCTGCTGATGATGCTGCTGATGATGCTGCTGATG 206

QY 216 TGGACCAAACTAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 275

DB 205 TCCTGCTGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 146

QY 276 TGCTTCCCAACACAGATCCTAAAAATGGTG 304

DB 145 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 117

RESULT 16

ABS15980/c

ID ABS15980 standard; DNA; 439 BP.

XX AC ABS15980;

XX 19-AUG-2002 (first entry)

XX Human genome-derived single exon probe ORF from lung SEQ ID No 15971.

XX Human; ds; single exon probe; asthma; lung cancer; COPD; IUD;

KW chronic obstructive pulmonary disease; interstitial lung disease;

KW familial idiopathic pulmonary fibrosis; neurofibromatosis;

KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;

KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;

KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;

KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;

KW primary ciliary dyskinesia; pulmonary hypertension;

XX hyaline membrane disease; open reading frame; ORF.

XX Homo sapiens.

XX WO200186003-A2.

XX 15-NOV-2001.

XX 30-JAN-2001; 2001WO-US00665.

XX 04-FEB-2000; 2000US-180312P.

XX 26-MAY-2000; 2000US-207456P.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-234687P.

PR 27-SEP-2000; 2000US-236359P.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2002-114183/15.
XX
XX Spatially-addressable set of single exon nucleic acid probes, used to
XX measure gene expression in human lung samples -
XX
XX Claim 4; SEQ ID No 15971; 634pp; English.
XX
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human lung comprising single exon nucleic acid probes having one of
XX 12614 nucleic acid sequences mentioned in the specification, or their
XX complements or the 12387 open reading frames derived from the 12614
XX probes. Also included are a microarray comprising the novel set of
XX probes; the novel set of probes which hybridize at high stringency to a
XX nucleic acid expressed in the human lung; measuring gene expression in a
XX sample derived from human lung, comprising (a) contacting the array with
XX a collection of detectably labeled nucleic acids derived from human lung
XX mRNA, and (b) measuring the label detectably bound to each probe of
XX the array; identifying exons in a eukaryotic genome, comprising
XX (a) algorithmically predicting at least one exon from genomic sequences
XX of the eukaryote; and (b) detecting specific hybridisation of detectably
XX labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
XX having a fragment identical to the predicted exon, the probe is included
XX in the above mentioned microarray; assigning exons to a single gene,
XX comprising (a) identifying exons from genomic sequence by the method
XX above and (b) measuring the expression of each of the exons in several
XX tissues and/or cell types using hybridisation to a single exon
XX microarrays having a probe with the exon, where a common pattern of
XX expression of the exons in the tissues and/or cell types indicates that
XX the exons should be assigned to a single gene; a peptide comprising one
XX of 12011 sequences, mentioned in the specification, or encoded by the
XX probes/open reading frames (ORF). The probes are used for gene
XX expression analysis, and for identifying exons in a gene, particularly
XX using human lung derived mRNA and for the study of lung diseases
XX such as asthma, lung cancer, chronic obstructive pulmonary disease
XX (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
XX fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
XX Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
XX haemorrhoidosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
XX pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
XX pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
XX and hyaline membrane disease. The present sequence is a single exon
XX probe open reading frame of the invention.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WPI at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 439 BP; 141 A; 169 C; 41 G; 88 T; 0 other;

Query Match 3.6%; Score 40.2; DB 24; Length 439;
Best Local Similarity 46.8%; Pred. No. 0.74;
Matches 126; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

QY 36 TGAAGGTGACAAAGCTGTTGTTAAACAGATGCTCAGTTCACAGATTAAGGAGGTAC 95
DB |||||
385 TGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 326
QY 96 AGCTTCGGTGAAGGTTGAGGCTGTTGCTGGTAAACCCCAACTGATGGAGCATATTGCTTA 155
DB |||||
325 TGCCGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 266
QY 156 TAAGATTGGTCCAGAAAGTTCAATTCTAGGATGGACATTTGCTGGTTCAGTTGCAACT 215
DB |||||
265 TGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 206

216 TGGACCAAAATGCTAGTACTGACTTGAAGGTTGGAGATACCGGTTTCGGTTTTCACGG 275
205 TCCTGCTGATGGTTTGGATGGTTATGATGATGATGATGATGATGATGATGATGATGATG 146
276 TGCTTCCCAACAGATCCTAAATAATGGTG 304
145 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 117

RESULT 17
AAD44410/c
ID AAD44410 standard; DNA; 1236 BP.
XX
XX AAD44410;
XX
XX 13-DEC-2002 (first entry)
XX Human huntington (htQ103) protein encoding DNA.
XX
XX Human; protein misfolding; Alzheimer's disease; AD; Parkinson's disease;
XX PD; Familial amyloid polyneuropathy; tauopathy; frontotemporal dementia;
XX Pick disease; lobar atrophy; trinucleotide disease; fragile-X syndrome;
XX Huntington's disease; spinocerebellar ataxia; SCA; myotonic dystrophy;
XX dentatorubral pallidoluysian atrophy; DRPLA; Creutzfeldt-Jacob disease;
XX CJD; prion disease; Gerstmann-Strausler-Scheinker disease; GSS; PFI;
XX fatal familia insomnia; mad cow disease; scrapie; kuru; anticonvulsant;
XX neurotropic; neuroprotective; cerebroprotective; htQ103 protein; gene; ds.
XX
XX Homo sapiens.
XX
XX OS
XX
XX Key
XX CDS 1..513
XX /location= "Human htQ103 protein"
XX /tag= a
XX /note= "No stop codon"
XX /partial
XX
XX WO200265136-A2.
XX
XX 22-AUG-2002.
XX
XX 15-FEB-2002; 2002WO-US04632.
XX
XX 15-FEB-2001; 2001US-269157P.
XX (UYCH-) UNIV CHICAGO.
XX
XX Lindquist S, Krobitsch S, Outeiro T;
XX
XX WPI; 2002-667026/71.
XX P-PSDB; AAE26650.
XX
XX Screening for therapeutic agents for protein misfolding disease, by
XX contacting a yeast cell with compound, that expresses misfolded disease
XX protein, and with a toxicity inducing agent, and evaluating cell for
XX viability -
XX
XX Disclosure; Page 87-88; 93pp; English.
XX
XX The present invention relates to novel screening methods for identifying
XX therapeutic agents for diseases associated with protein misfolding. The
XX method involves contacting a yeast cell with a candidate compound, where
XX the yeast cell expresses a polypeptide comprising a misfolded disease
XX protein, contacting the yeast cell with a toxicity inducing agent and
XX evaluating the yeast cell for viability, where the viability indicates
XX the candidate compound is a candidate therapeutic agent. The method is
XX useful to screen for therapeutic agents for diseases associated with
XX protein misfolding such as Alzheimer's disease (AD), Parkinson's disease
XX (PD), Familial amyloid polyneuropathy, tauopathies (e.g. Pick disease,
XX lobar atrophy, frontotemporal dementia) or trinucleotide diseases (e.g.
XX Huntington's disease, spinocerebellar ataxia (SCA), fragile-X syndrome,
XX myotonic dystrophy, dentatorubral pallidoluysian atrophy (DRPLA) and
XX prion diseases (e.g. Creutzfeldt-Jacob disease (CJD), fatal familia

XX Disclosure; SEQ ID NO 22285; 3071pp + Sequence Listing; English.

XX AA54951 to AA64702 encode the human immune/haematopoietic antigen (I)

XX amino acid sequences given in AA82170 to AA81921. (I) have cytostatic

XX activity, and can be used in gene therapy and vaccine production. (I)

XX proteins and polynucleotides may be used in the prevention, diagnosis and

XX treatment of diseases associated with inappropriate (I) expression. For

XX example, they may be used to treat disorders associated with decreased

XX expression by rectifying mutations or deletions in a patient's genome

XX that affect the activity of (I) by expressing inactive proteins or to

XX supplement the patient's own production of (I). Additionally, (I)

XX polynucleotides may be used to produce the secreted (I), by inserting

XX the nucleic acids into a host cell and culturing the cell to express the

XX protein. (I) proteins and polynucleotides may be used to prevent,

XX diagnose and treat immune/haematopoietic-related diseases, especially

XX cancers and cancer metastases of haematopoietic-derived cells. AA64703

XX to AA87694 represent human immune/haematopoietic antigen genomic

XX sequences from the present invention. AA54942 to AA54950 and AA82169

XX represent sequences used in the exemplification of the present invention.

XX SQ Sequence 4027 BP; 1247 A; 875 C; 797 G; 1108 T; 0 other;

Query Match 3.6%; Score 39.6; DB 22; Length 4027;

Best Local Similarity 44.0%; Pred. No. 2.7;

Matches 168; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

Qy 639 TTATGGTGCTGATGCTTTGACTATCATGATGCGAGCGGTATTGAGCAGATCAATC 698

Db 987 TTCAGCGATTCTCTTGCTCAGGCTCCCAAGTAGCTGGGATTACAGGTGTGCCACCA 1046

Qy 699 GAAGTATCCAACTCCCAACATGTTATTGACGCTGTGGAGCGAAGATAGTATCCCGA 758

Db 1047 CACTCAGCTAATTTTGTATTTAGTAGCGCAGGGTTTGGCATGTTACACGAGCTGG 1106

Qy 759 GGCTTATAAGTCACAGCAGATGCTTACCTGCGCACATATTAGAAGTGGTTCCAATGAC 818

Db 1107 TCTCGAATCTCGACCTCAGGTGATCGCCCTGCGCTCCCAAAAGTGTGGGATTAC 1166

Qy 819 CATTTGAAGCATTCCTGAGGAATCAGAAAGATAAGTTAAATGATATTAATTACTTGT 878

Db 1167 AGGCGTAGCCACCAAAACCCAGCTTAATCAATTTGTTCTAATAAAGAAATGTGAAC 1226

Qy 879 GTATCGTGCTCTGCTCAAGAAATTTCTATTGGGTGCAACAGATTTCTGCTAGTCCAGA 938

Db 1227 GAGAAACCATTTAAGGAAGTAATCAATTTAGTCAACGGGACTTCTGAGATTTCACT 1286

Qy 939 ATATCATGAAGCCACAGTTAAATTCGTTAAGTTTATAAATCCACCTTAACAAACGGTGA 998

Db 1287 TACTTAAAGTGAATAATGTATACATTTTATTCTTCAAAACACAGAAACCTGGGTTA 1346

Qy 999 TATCCATCATATGATATTTAA 1020

Db 1347 GATGGTTAAATGATGGTTAA 1368

RESULT 20

ABL55644/c

ID ABL55644 standard; DNA; 50000 BP.

XX AC

XX ABL55644;

XX 01-JUL-2002 (first entry)

XX DE AmEPV genome fragment#2.

XX AmEPV; gene therapy; viral vector; chromosome mapping; gene mapping;

XX genetic deficiency disorder; da.

XX Amsacta moorei entomopoxvirus.

XX WO200212526-A2.

XX

PD 14-FEB-2002.

XX 10-AUG-2001; 2001WO-US25287.

XX 10-AUG-2000; 2000US-224479P.

PR 14-SEP-2000; 2000US-0662254.

XX (UYFL) UNIV FLORIDA.

XX Moyer RW, Li Y, Bawden AL;

PI WPI; 2002-227161/28.

XX Novel recombinant entomopox virus vector useful for delivering

PT polynucleotide encoding protein to vertebrate cell, comprises

PT polynucleotide encoding protein operably linked with heterologous

PT promoter sequence -

XX Disclosure; Page 150-175; 326pp; English.

XX The invention relates to a recombinant entomopox virus (EPV) vector,

CC comprising a polynucleotide encoding a protein operably linked with a

CC heterologous promoter sequence. The invention also concerns methods for

CC providing gene therapy for genetic deficiency disorders. Vectors of the

CC invention are useful for delivering a polynucleotide encoding a protein

CC to a vertebrate cell preferably a mammalian cell, such as a human cell.

CC The vector is introduced into the vertebrate cell by infection in a viral

CC particle, or by transfection, transduction, or injection either in vitro

CC or in vivo. The vector is useful for the delivery and expression of

CC biologically useful proteins in gene therapy protocols, and for

CC delivering large DNA segments for engineering of vertebrate cells.

CC Polynucleotides of the invention have applications in techniques such as

CC their use as insertion sites for foreign genes of interest, hybridisation

CC probes, for chromosome and gene mapping, in PCR technologies, and in the

CC production of sense or antisense nucleic acids. Vectors of the invention

CC provide for stable integration and expression of heterologous DNA in host

CC cells, and are adapted for accepting large heterologous polynucleotide

CC inserts which can be delivered in an infected or transformed cell and

CC expressed in a stable fraction. The current sequence represents a

CC fragment of the genome of the genus B entomopoxvirus from amsacta moorei

CC (AmEPV).

XX SQ Sequence 50000 BP; 20514 A; 4505 C; 4614 G; 20367 T; 0 other;

Query Match 3.6%; Score 39.6; DB 24; Length 50000;

Best Local Similarity 49.5%; Pred. No. 7.5;

Matches 102; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

Qy 889 TCTGGTCAAGAAATTCATTGGGTGCAACAGATTTCTCTAGTCCAGAAATATCATGAA 948

Db 1619 TATTATGAAAAAATTTATTATATATAAAAAACCAATTTCTATGTATAGCAAAATATCATGAA 1560

Qy 949 GCCACAGTTAAATTCGTTAAGTTTATAATCCACACCTTAACACGGTGTATCCATCAT 1008

Db 1559 GAAAAAGATATTTATACATAGATATATAAGATATAATCATTTATGATGATATTTCCCTAAAGAA 1500

Qy 1009 ATGAATATTAAAGTTTTCAGCAACGGCTTAGATGATGTCCTCAGCTCTCACTGAAGGTATA 1068

Db 1499 AATATAAAAAAATTTCTTTTATGATATATTTAATAAATAGCAGATATATTTTGAATAATTT 1440

Qy 1069 AAAGAAGGTAAAAACAAAAATGTTAA 1094

Db 1439 AAAATTAAAAAATAATATGTTGATTA 1414

RESULT 21

AAT91162

ID AAT91162 standard; cDNA; 772 BP.

XX AC AAT91162;

XX 27-MAR-1998 (first entry)

XX

heterologous promoter sequence. The invention also concerns methods for providing gene therapy for genetic deficiency disorders. Vectors of the invention are useful for delivering a polynucleotide encoding a protein to a vertebrate cell, preferably a mammalian cell, such as a human cell. The vector is introduced into the vertebrate cell by infection in a viral particle, or by transfection, transduction, or injection either in vitro or in vivo. The vector is useful for the delivery and expression of biologically useful proteins in gene therapy protocols, and for delivering large DNA segments for engineering of vertebrate cells. Polynucleotides of the invention have applications in techniques such as their use as insertion sites for foreign genes of interest; hybridisation probes, for chromosome and gene mapping, in PCR technologies, and in the production of sense or antisense nucleic acids. Vectors of the invention provide for stable integration and expression of heterologous DNA in host cells, and are adapted for accepting large heterologous polynucleotide inserts which can be delivered in an infected or transformed cell and expressed in a stable fraction. The current sequence represents a fragment of the genome of the genus B entomopoxvirus from *Amblyopoda moorei* (AmEPV).

Sequence 5000 BP; 19857 A; 4373 C; 4222 G; 21548 T; 0 other;

Query Match 3.4%; Score 38; DB 24; Length 50000;

Best Local Similarity 47.5%; Fred. No. 21;

Matches 113; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

Qy	859	AAATTTGATATTACTTTGTTGTCATCTGTCAGAAATTCATTGGTGCAACA	918
Db	45093	AAAGGTATAGTACCAATGATATTGTTACTTTTATTATAAATAAATAATATCC	45152
Qy	919	AGATTTCTGCTAGTCCAGAAATATCATGAAGCCACAGTTAAATTCGTTAAGTTTATAAAT	978
Db	45153	AAATAAAGCTCCCATGTTTAAAAATAATAAAGTCACAGTTATATATAAATAATATAT	45212
Qy	979	CCACACCTTAACAACGGTGATATCCATCATATGATATATAAGTTTTCAGCAACGGCTTA	1038
Db	45213	TTTTTCATTATAGTATTTTATATAAAAAATAATAATAAATAAATGTTTAAACACAGATTTA	45272
Qy	1039	GATGATGTCACGCTCTCACTGAGGTATATAAAGAGGTAAAAACAAAAATGTTAAGT	1096
Db	45273	ACTATGAGAGATATCAGAGCTGCTAATAAATAATAAATAAATAATATCTTGTAAT	45330

Search completed: January 10, 2004, 06:40:13
Job time : 403 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2004, 06:26:03 ; Search time 88 Seconds
(without alignments)
5582.493 Million cell updates/sec

Title: US-10-081-644-1
Perfect score: 1113
Sequence: 1 atgcagttcccaaccactca.....agatgttcaggattataa 1113

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : Issued Patents NA.*

1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*

2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*

3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*

4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*

5: /cgn2_6/ptodata/2/ina/PCTUS COMB.seq.*

6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37.6	3.4	8654	4	US-08-961-527-98
c	36.6	3.3	1664976	4	US-08-916-421B-1
3	36.2	3.3	3255	4	US-09-601-198-108
4	36.2	3.3	580073	4	US-08-545-528D-1
5	36	3.2	1830121	4	US-09-557-884-1
6	36	3.2	1830121	4	US-09-643-990A-1
7	35.8	3.2	840	3	US-09-042-771-1
8	35.6	3.2	849	4	US-09-107-532A-403
c	35.6	3.2	3312	4	US-09-601-198-58
10	35.6	3.2	3600	3	US-08-855-910-7
11	35.6	3.2	3792	2	US-08-992-334-1
12	35.6	3.2	3792	3	US-08-302-752-1
13	35.6	3.2	5234	3	US-08-992-334-2
14	35.6	3.2	5234	3	US-08-302-752-2
15	35.6	3.2	6722	2	US-08-992-334-3
16	35.6	3.2	6722	3	US-08-302-752-3
17	35.2	3.2	1434	4	US-09-134-001C-650
18	35	3.1	927	4	US-09-328-352-2636
19	35	3.1	4860	4	US-09-328-352-3221
20	34.8	3.1	360	4	US-09-702-705-1408
21	34.8	3.1	360	4	US-09-736-457-1408
22	34.8	3.1	1044	4	US-09-134-001C-461
c	34.8	3.1	1053	4	US-09-107-532A-3324
24	34.8	3.1	1347	3	US-09/622
25	34.8	3.1	1347	4	US-08-165-922A-11
c	34.8	3.1	1374	4	US-09-601-198-131
27	34.8	3.1	4016	4	US-09-173-053-3

ALIGNMENTS

RESULT 1

US-08-961-527-98

; Sequence 98, Application US/08961527

; Patent No. 6420135

; GENERAL INFORMATION:

Sequence 209, App
Sequence 209, App
Sequence 29, Appl
Sequence 3644, Ap
Sequence 1, Appl
Sequence 1, Appl
Sequence 14, Appl
Sequence 1, Appl
Sequence 2004, Ap
Sequence 1, Appl
Sequence 1, Appl
Sequence 3, Appl
Sequence 1, Appl
Sequence 2, Appl
Sequence 60, Appl
Sequence 1, Appl
Sequence 382, App
Sequence 2274, Ap
Sequence 21, Appl
Sequence 12, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 218, App
Sequence 1, Appl
Sequence 3, Appl
Sequence 3731, Ap
Sequence 26, Appl
Sequence 61, Appl
Sequence 157, App
Sequence 157, App
Sequence 157, App
Sequence 157, App
Sequence 278, App
Sequence 278, App
Sequence 278, App
Sequence 3001, Ap
Sequence 5, Appl
Patent No. 5231168
Sequence 50, Appl
Sequence 212, App
Sequence 1334, Ap
Sequence 1062, Ap
Sequence 7, Appl
Sequence 55, Appl
Sequence 3, Appl
Sequence 9, Appl
Sequence 1, Appl
Sequence 48, Appl
Sequence 48, Appl
Sequence 48, Appl
Sequence 48, Appl
Sequence 30, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 124, App
Sequence 224, App
Sequence 9, Appl
Sequence 1, Appl
Sequence 69, Appl
Sequence 77, Appl
Sequence 95, Appl
Sequence 5, Appl
Sequence 5, Appl

APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 98:
SEQUENCE CHARACTERISTICS:
LENGTH: 8654 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-98

Query Match 3.4%; Score 37.6; DB 4; Length 8654;
Best Local Similarity 54.3%; Pred. No. 1.8;
Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
Qy 788 CTGCCACATTATTAGAGTGGTCCCAATGACCATTTGATCGTCATCTGTCAGAAATCTAT 907
Db 6452 CTTTCACATCTGTAAGTGTGATGCGAATTTGGATGATCTATTGAAGTGGAAATCGTG 6511
Qy 848 AAGATATGTTAAATGATATTACTTTGTTGTTATCGTCATCTGTCAGAAATCTAT 907
Db 6512 AAGATGATATCAAGATGGATACCTTCGTTTCAGTGGTGGCGGACAAACGTCATA 6571
Qy 908 TGGGTGCAACAAGATTTCTCT 927
Db 6572 AGGTTTCAACAGGTGTACGT 6591

RESULT 2
US-08-916-421B-1/c
Sequence 1, Application US/08916421B
Patent No. 6503729
GENERAL INFORMATION:
APPLICANT: Bult et al.
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus jannaschii
TITLE OF INVENTION: jannaschii
FILE REFERENCE: PB275
CURRENT APPLICATION NUMBER: US/08/916,421B
CURRENT FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/024,428
PRIOR FILING DATE: 1996-08-22
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1664976
TYPE: DNA

ORGANISM: Methanococcus jannaschii
FEATURE:
NAME/KEY: misc feature
LOCATION: (28222)..(28222)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (28257)..(28258)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84773)..(84773)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84808)..(84808)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84812)..(84812)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98120)..(98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (103998)..(103998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (231980)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234814)..(234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (319226)..(319226)

```
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (559167)..(559167)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (559241)..(559241)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (600992)..(600992)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (622708)..(622708)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (657081)..(657081)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (657203)..(657203)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (674435)..(674435)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (682442)..(682442)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (713652)..(713652)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (741684)..(741684)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (779455)..(779455)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (779676)..(779676)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (855539)..(855539)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (871619)..(871619)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1084830)..(1084830)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1096846)..(1096846)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1119881)..(1119881)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1130881)..(1130881)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1310988)..(1310988)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1313224)..(1313224)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1349473)..(1349473)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1349491)..(1349491)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1470091)..(1470091)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1569020)..(1569020)
/ OTHER INFORMATION: n equals a, t, c, or g
```

```
/ NAME/KEY: misc feature
/ LOCATION: (1602912)..(1602912)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1603734)..(1603734)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1637998)..(1637998)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1664855)..(1664855)
/ OTHER INFORMATION: n equals a, t, c, or g
/ US-08-916-421B-1

Query Match      3.3%; Score 36.6; DB 4; Length 1664976;
Best Local Similarity 49.7%; Pred. No. 27;
Matches 93; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 723 TATTGACGCTGTGGGAAGCAAGATAGTATCCCCGAGGCTATAAAGTCACAGCAGATAG 782
Db 1311832 TTTTGAATCTGATAAAGTAGGAGAGATTTTAAATAAATAGATAAAGATACTCTAGTTAT 1311773

QY 783 TCTACCTGCCACATTATTAGAAAGTGTTCACATGACCATTTGAAGCATTCCTGAAGAAAT 842
Db 1311772 TATCACTGTTCAACAATCGATTTGAAAGTTATTGAACTTTGAAACTTTTAAAAAGTTAT 1311713

QY 843 CAGAAAGATAATGTTAAAAATTGATATTACTTTGTTGATCGTGCACTCGGTCAAGAAAT 902
Db 1311712 TAATCAGCTAAATAAATAAAGAGCTATTGCTTTGTTGATGATGCTCTGGAGCGAGAGT 1311653

QY 903 TCTATTG 909
Db 1311652 TAGGTTG 1311646

RESULT 3
US-09-601-198-108/c
; Sequence 108, Application US/09601198
; Patent No. 6531583
; GENERAL INFORMATION:
; APPLICANT: Caspell, Gail H.
; APPLICANT: Chen, Ellison Y.
; APPLICANT: Glass, Jennifer S.
; APPLICANT: Glass, John I.
; APPLICANT: Heiner, Cheryl R.
; APPLICANT: Leftowitz, Elliot
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA
; TITLE OF INVENTION: UREALYTICUM
; FILE REFERENCE: UAB-13452/22
; CURRENT APPLICATION NUMBER: US/09/601,198
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/073,189
; PRIOR FILING DATE: 1998-01-30
; NUMBER OF SEQ ID NOS: 181
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 108
; LENGTH: 3255
; TYPE: DNA
; ORGANISM: Ureaplasma urealyticum
/ US-09-601-198-108

Query Match      3.3%; Score 36.2; DB 4; Length 3255;
Best Local Similarity 45.8%; Pred. No. 2,9;
Matches 125; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

QY 829 ATTCCTGAAGAAATCAGAAAAGATAATGTTAAATTTGATATTACTTTGTTGATCGTCA 888
Db 3225 ATAGGTGTTGTATCAACAATAATTTCTTGGTTCAATAAATATATCATTTCTAATTGTTCT 3166

QY 889 TCTGGTCAAGAAATTTCTATTGGGTGCAACAAGATTTTCCTGCTAGTCCAGATATCATGAA 948
Db 3165 TATAACAAAAAACACATTTTCGTCCTTCTACAATTTACGAATCAAGATATAAATAATTTAGAC 3106
```

QY 949 GCCACAGTTAAATCGTTAAGTTTATAAAATCCACACCTTAACACGGTGATATCCATCAT 1008
Db 3105 ATAATAGCAAAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3046
QY 1009 ATGAATATTAAGTTTTCAGCAACGGCTTAGATGATGTCCTCCAGCTCTCACTGAAGGTATA 1068
Db 3045 ATTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2986
QY 1069 AAAGAGGTAAAAACAAATGTTTAAGTATGTT 1101
Db 2985 AAAGTTATTATTGAACAAATGCTCAATATTTT 2953

RESULT 4
US-08-545-528D-1
; Sequence 1, Application US/08545528D
; Patent No. 6537773
; GENERAL INFORMATION:
; APPLICANT: Fraser et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment
; Patent No. 6537773
; FILE REFERENCE: P0193P1
; CURRENT APPLICATION NUMBER: US/08/545,528D
; PRIOR FILING DATE: 1995-10-19
; PRIOR APPLICATION NUMBER: US 08/488,018
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/473,545
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 580073
; TYPE: DNA
; ORGANISM: Mycoplasma genitalium
US-08-545-528D-1

Query Match 3.3%; Score 36.2; DB 4; Length 580073;
Best Local Similarity 56.2%; Pred. No. 23;
Matches 68; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
QY 792 CACATTATTAGAAGTGGTTCACATGACCAATGAAAGCATTCCTGAAAGAAATCAGAAAAGA 851
Db 464202 CAGATTATTGGAATGGCTACATTGATTATGAACAAATATATAAGAAAAAGTAAAGC 464261
QY 852 TAATGTTAAATGATATTACTTTGTTGATCTGATCTGGTCAAGAAATCTATTGGG 911
Db 464262 TATTGTGCAATCTTTTGACTAGCTGATATTTTACTATTGAAAAATAACTGTTTGGC 464321

QY 912 T 912
Db 464322 T 464322

RESULT 5
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850
; COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
; QY 912 T 912
Db 464322 T 464322

COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/557,884
FILING DATE: 25-Apr-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: JUN-5-1995
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1
Query Match 3.2%; Score 36; DB 4; Length 1830121;
Best Local Similarity 51.2%; Pred. No. 39;
Matches 84; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 28 GTCATCATTTGAAGGTGACAAAGCTGTTGTTAAAAACAGATGCTCTCAGTTCAGAAATTAAAG 87
Db 844860 GTGTTGTTGAAGGTATCAACATCATTAATAACATGAACCAACAGTTCCTGCATTAGCA 844919
QY 88 GAGGGTACAGCTTGGTGAAGGTGGAGCTGTTGGTGAACCAACTGATTGGAAGCAT 147
Db 844920 AAGGCTGGTGGTGTGTAAGAAAGAAAGCTGCAATTGACGTTCAAAATGTCGATTTTC 844979
QY 148 ATTGCTTATAAGATTGGTCCAGAAAGTTCAATCTAGGATGTA 191
Db 844980 AATCTTAAACAAATAAAGCTGACCGTGTAGGTTTGTAGATTGCA 845023

RESULT 6
US-09-643-990A-1
; Sequence 1, Application US/09643990A
; Patent No. 6528289
; GENERAL INFORMATION:
; APPLICANT: Robert D. Fleischmann
Mark D. Adams
Owen White
Hamilton O. Smith
J. Craig Venter
; TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville,
STATE: MD
COUNTRY: USA
ZIP: 20850
; COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,990A
FILING DATE: 23-Aug-2000
CLASSIFICATION: <Unknown>

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: 1995-06-07
; APPLICATION NUMBER: 08/426,787
; FILING DATE: 1995-04-21
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB186P1C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-610-5790
; TELEFAX: 310-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1
Query Match 3.2%; Score 36; DB 4; Length 1830121;
Best Local Similarity 51.2%; Pred. No. 39;
Matches 84; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
Qy 28 GTCATCATTTGAAGGTGACAAAGCTGTTGTTTAAACACAGATGTCAGTTCAGAAATTAAG 87
Db 844860 GTGTTGTTGAAGGTATCAACATCATTAACATGAAACCAACAGTTCCTGCATTAGGA 844919
Qy 88 GAGGTACAGCTTGGTGAAGTTGAGCGTGTCTGCTGTTACCACTGATTTGAAGCAT 147
Db 844920 AAGCGTGGTGGTTGGTGAAGAAAGAGCTGCAATTCAGCGTTCATAATGTTGCGATTTTC 844979
Qy 148 ATTGCTTATAGATTGCTCAGAGGTTCAATTCATGAGTGA 191
Db 844980 ATCTTAACAATAAAGCTGACCGTGTAGTTTAGATTGCA 845023
RESULT 7
US-09-042-771-1
; Sequence 1, Application US/09042771
; Patent No. 6080729
; GENERAL INFORMATION:
; APPLICANT: Jaworski, Deborah
; APPLICANT: Wang, Min
; APPLICANT: Shilling, Lisa K.
; APPLICANT: Lawlor, Elizabeth
; TITLE OF INVENTION: sp00J2
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/042,771
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Falk, Stephen T
; REGISTRATION NUMBER: 36,795
; REFERENCE/DOCKET NUMBER: GM10135
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:
; TELEFAX:
; INFORMATION FOR SEQ ID NO: 403:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 849 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: 1995-06-07
; APPLICATION NUMBER: 08/426,787
; FILING DATE: 1995-04-21
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB186P1C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-610-5790
; TELEFAX: 310-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 840 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-09-042-771-1
; TELEPHONE: 215-994-2488
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 840 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-09-042-771-1
Query Match 3.2%; Score 35.8; DB 3; Length 840;
Best Local Similarity 51.6%; Pred. No. 2.1;
Matches 82; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
Qy 281 CCCAAACAGATCCTAAAAATGGTGCAATTTGCTCAATATGCCAGGGTTTATCCACCTTGT 340
Db 290 CTCAGCAGACGTTATTATTCGTGATATGATGATGATGATGATGATGATGATGATGATGAT 349
Qy 341 TTTACAAGAGTAACTTAACCTCACTCAACTGCTGATGAAATTTCTGAAGGCCCTGTGAAGA 400
Db 350 TTGAGAATATTCAAGAGAGAAATTTGCTGCTGTTGTTGAAGAGCGGAAAGCCTATAAGAAAT 409
Qy 401 ACTTCGAATCTGCTGCATCATTTGCCAGTTTCCTGTTGACAA 439
Db 410 TATTGGAATTTGTTGATACAAACGCAAAAGTGAATTGGCAA 448
RESULT 8
US-09-107-532A-403
; Sequence 403, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A. Doucette-Stamm and David Bueh
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 403:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 849 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
```

```
;
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...849
; SEQUENCE DESCRIPTION: SEQ ID NO: 403:
US-09-107-532A-403

Query Match          3.2%; Score 35.6; DB 4; Length 849;
Best Local Similarity 46.7%; Pred. No. 2.4;
Matches 113; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

Qy 867 TATTACTTGTGTCATCGTCGTCATCTGCTCAAGAAATCTATTGGTGCAACAAGATTTC 926
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 543 TGTTCCAGTCGTTGATGTTTAACTTAAGTTGTTTCTAGTATTGAAACAAGTAAC 602
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 927 TGCTAGTCCAGATATCATGAAGCCACGAGTTAAATTCGTTAAAGTTTATAATCCACCT 986
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 603 TGCTGATCAAGTAACGAAGCAATCAAAAAACACACAGTTGACAATCCTTCATTGGGTA 662
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 987 TAACAAGGTCATATCCATCATATGAATATAAGTTTTCAGCAACGCTTAGATGATGT 1046
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 663 TGATGACCGTCAAAATCGTTCTAGCGATGTCATGGAACCACTCAAGTTTCGATCTTGA 722
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1047 CCCAGCTCTCACTGAAGGTATATAAGAGGATGAAGGATGAAGGATGAAGGATGAAGG 1106
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 723 CCCAAGCAACAGAGTAACAACTGCTGCTGATTTCCAACTGTTAAACTGTTGCATG 782
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1107 GT 1108
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 783 GT 784

RESULT 9
US-09-601-198-58/c
; Sequence 58, Application US/09601198
; Patent No. 6531583
; GENERAL INFORMATION:
; APPLICANT: Cassell, Gail H.
; APPLICANT: Chen, Ellison Y.
; APPLICANT: Glass, Jennifer S.
; APPLICANT: Glass, John I.
; APPLICANT: Heiner, Cheryl R.
; APPLICANT: Lefkowitz, Elliot
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA
; FILE REFERENCE: UAB-13452/22
; CURRENT APPLICATION NUMBER: US/09/601.198
; CURRENT FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/073,189
; PRIOR FILING DATE: 1998-01-30
; NUMBER OF SEQ ID NOS: 181
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 58
; LENGTH: 3312
; TYPE: DNA
; ORGANISM: Ureaplasma urealyticum
US-09-601-198-58

Query Match          3.2%; Score 35.6; DB 4; Length 3312;
Best Local Similarity 48.1%; Pred. No. 4.2;
Matches 101; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

Qy 894 TCAAGAAATTCATTATGGGTGCAACAAGATTTCCTGCTAGTCCAGATATCATGAAGCCAC 953
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2893 TAATGAAGATTTAAATTTACAAACCATCACTCTCTAATACACCAAGATTGATCATTTAGA 2834
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 954 AGTTAAATTCGTTAAAGTTTATAATCCACACCTTAACAAGGTCATATCCATCATATGAA 1013
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2833 AATTAAATTAATTCGTTTACAAATATAAACAACCAAGTTGATATAAAGATTAAGTT 2774
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1014 TATTAAAGTTTTCAGCAACCGCTTAGATGATGTGCCAGCTCTCACTGAAGGTATAAAGA 1073
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db 2773 AAAAAAGAGTTTCTTTTAAAGAGCATGCTATTCATATTTTAGTTTAAAAAATATCAAAA 2714
Qy 1074 AGTAAAAACAAAAATGTTAAGTATGCTGC 1103
Db 2713 AACTAATCAACCAATCATTACTAGTAGTC 2684

RESULT 10
US-08-855-910-7
; Sequence 7, Application US/08855910
; Patent No. 6221640
; GENERAL INFORMATION:
; APPLICANT: Tao, Jianehi
; APPLICANT: Sasanfar, Mandana
; APPLICANT: Gallant, Paul L.
; APPLICANT: Shen, Xiaoyu
; APPLICANT: Avruch, Anthony S.
; APPLICANT: Yu, Russell V.
; APPLICANT: Nair, Shamila
; TITLE OF INVENTION: ENTEROCOCCAL AMINOACYL-tRNA SYNTHETASE
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACIDS AND STRAINS COMPRISING SAME
; NUMBER OF SEQUENCES: 71
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/855,910
; FILING DATE: 14-MAY-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: CPI95-08
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781) 861-6240
; TELEFAX: (781) 861-9540
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3600 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 85..1128
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1139..3559
US-08-855-910-7

Query Match          3.2%; Score 35.6; DB 3; Length 3600;
Best Local Similarity 51.9%; Pred. No. 4.4;
Matches 80; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

Qy 708 AAACCTGCAACATGTTATTGAGCTGTGGAGAGCGAAGATAGTATCCCCGAGCCTATAA 767
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3223 AAACTGCAACATGTTAGTAGCCACAGACGAGCGGGTTACTTATGAAAGCAGTTTCTAA 3282
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 768 AGTCACAGCAGATAGTCTACCTGCCACATATTAGAGTGGTTCCTCAATGCCATTCGAAG 827
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3283 ATTCCAGCAGTTTCTCGAGATATTGCTTTATTAGTCGATGAACAGTTTACCANTCAAG 3342
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 828 CATTCCTCGAAGAAATCAGAAAAAGATAATGTTAAA 861
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```


Db 3343 ACTAGTAAACTATTTTCAGATAACCGAGGTAAA 3376

RESULT 11

US-08-992-334-1
; Sequence 1, Application US/08992334
; Patent No. 5919678
; GENERAL INFORMATION:
; APPLICANT: Gruss, Alexandra
; APPLICANT: Maguin, Emmanuelle
; TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
; TITLE OF INVENTION: PLASMID
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christie Parker & Hale, LLP
; STREET: 350 West Colorado Boulevard, Suite 500
; CITY: Pasadena
; STATE: California
; COUNTRY: United States
; ZIP: 91105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/992,334
; FILING DATE: 17-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,752
; FILING DATE: 24-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR93/00248
; FILING DATE: 12-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR FR92/03034
; FILING DATE: 13-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Prout, D. Bruce
; REGISTRATION NUMBER: 20958
; REFERENCE/DOCKET NUMBER: C93:31779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (626) 795-9900
; TELEFAX: (626) 577-8800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3792 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
; IMMEDIATE SOURCE:
; CLONE: pg+host4

US-08-992-334-1

Query Match 3.2%; Score 35.6; DB 2; Length 3792;
Best Local Similarity 47.3%; Pred. No. 4.5;
Matches 107; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

Qy 820 ATTGAAGCATTCCTGGAAGAAATCTAATGCGTCAACAGATTTCTCTGCTAGTCCAGAA 879
Db 2229 ATTGACGAATCTGTAACTAAGAGCGTTAGGAACAGATTAAGCGAAAATTTGGG 2288
Qy 880 TATCGTGCATCTGGTCAAGAAATCTAATGCGTCAACAGATTTCTCTGCTAGTCCAGAA 939
Db 2289 AATAGTTTCAGTTGCTCATGTTGAGTACTTCAATATATCAAGGTTTCATATGATATTG 2348
Qy 940 TATCATGAAGCCACAGTTAAATTCGTTAAATTTTAAATCCACACCTTAAACACGGTGAT 999
Db 2349 ACTCATGAATCAAGGACGCTATTGCTAAGATAAATATATATACGACAAAAGATATT 2408

Qy 1000 ATCCATCATATGAATATTAAAGTTTTCAGCAACGGCTTAGATGATG 1045
Db 2409 TTGAACATTAATGATTTTGATATTGACCGCTATATAACACTTGATG 2454

RESULT 12

US-08-302-752-1
; Sequence 1, Application US/08302752
; Patent No. 6025190
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: THERMOSENSIBLE PLASMID
; NUMBER OF SEQUENCES: 3
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,752
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9203034
; FILING DATE: 13-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO FR/93/00248
; FILING DATE: 12-MAR-1993
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3792 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-302-752-1

Query Match 3.2%; Score 35.6; DB 3; Length 3792;
Best Local Similarity 47.3%; Pred. No. 4.5;
Matches 107; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

Qy 820 ATTGAAGCATTCCTGGAAGAAATCTAAGAAATGATAATGTTAAATTTGATATTACTTTGTTG 879
Db 2229 ATTGACGAATCTGTAACTAAGAGCGTTAGGAACAGATTAAGCGAAAATTTGGG 2288
Qy 880 TATCGTGCATCTGGTCAAGAAATCTAATGCGTCAACAGATTTCTCTGCTAGTCCAGAA 939
Db 2289 AATAGTTTCAGTTGCTCATGTTGAGTACTTCAATATATCAAGGTTTCATATGATATTG 2348
Qy 940 TATCATGAAGCCACAGTTAAATTCGTTAAATTTTAAATCCACACCTTAAACACGGTGAT 999
Db 2349 ACTCATGAATCAAGGACGCTATTGCTAAGATAAATATATATACGACAAAAGATATT 2408
Qy 1000 ATCCATCATATGAATATTAAAGTTTTCAGCAACGGCTTAGATGATG 1045
Db 2409 TTGAACATTAATGATTTTGATATTGACCGCTATATAACACTTGATG 2454

RESULT 13

US-08-992-334-2
; Sequence 2, Application US/08992334
; Patent No. 5919678
; GENERAL INFORMATION:
; APPLICANT: Gruss, Alexandra
; APPLICANT: Maguin, Emmanuelle
; TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
; TITLE OF INVENTION: PLASMID
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christie Parker & Hale, LLP
; STREET: 350 West Colorado Boulevard, Suite 500
; CITY: Pasadena

ATTORNEY/AGENT INFORMATION:
NAME: Prout, D. Bruce
REGISTRATION NUMBER: 20958
REFERENCE/DOCKET NUMBER: C93:31779
TELECOMMUNICATION INFORMATION:
TELEPHONE: (626) 795-9900
TELEFAX: (626) 577-8800
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 6722 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
US-08-992-334-3

Query Match 3.2%; Score 35.6; DB 2; Length 6722;
Best Local Similarity 47.3%; Pred. No. 5.6;
Matches 107; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 820 ATTGAAGCATTCCTGAGAAATCAGAAAAAGATAATGTTAAAAATTGATATTACTTTGTTG 879
DB 5159 ATTGCACGAATCCTGTAACAATAGAAAGCGTTAGGAACAAGATTAAAGGAAAAATTTGGG 5218
QY 880 TATCGTGCATCTGGTCAAGAAATTCCTATTGGGGTGCACAGATTTTCCTCTAGTCCAGAA 939
DB 5219 AATAGTTCAGTTGCTCATCTTGAGATACCTTGATTTATATCAAGGTTTCATATGAATATTG 5278
QY 940 TATCATGAAGCCACAGTTAAATTCGTTAAAGTTTATAAAATCCACACCTTAAACACGGTGAT 999
DB 5279 ACTCATGAATCAAGGACGCTATTGCTAAGAATAAACAATATATACGACAAAAAGATATT 5338
QY 1000 ATCCATCATATGAATATTAAAGTTTTCAGCAACGGCTTAGATGATG 1045
DB 5339 TTGAACATTAAATGATTTTGATATTGACCGCTATATAACACTTGATG 5384

Query Match 3.2%; Score 35.6; DB 3; Length 6722;
Best Local Similarity 47.3%; Pred. No. 5.6;
Matches 107; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 820 ATTGAAGCATTCCTGAGAAATCAGAAAAAGATAATGTTAAAAATTGATATTACTTTGTTG 879
DB 5159 ATTGCACGAATCCTGTAACAATAGAAAGCGTTAGGAACAAGATTAAAGGAAAAATTTGGG 5218
QY 880 TATCGTGCATCTGGTCAAGAAATTCCTATTGGGGTGCACAGATTTTCCTCTAGTCCAGAA 939
DB 5219 AATAGTTCAGTTGCTCATCTTGAGATACCTTGATTTATATCAAGGTTTCATATGAATATTG 5278
QY 940 TATCATGAAGCCACAGTTAAATTCGTTAAAGTTTATAAAATCCACACCTTAAACACGGTGAT 999
DB 5279 ACTCATGAATCAAGGACGCTATTGCTAAGAATAAACAATATATACGACAAAAAGATATT 5338
QY 1000 ATCCATCATATGAATATTAAAGTTTTCAGCAACGGCTTAGATGATG 1045
DB 5339 TTGAACATTAAATGATTTTGATATTGACCGCTATATAACACTTGATG 5384

QY 820 ATTGAAGCATTCCTGAGAAATCAGAAAAAGATAATGTTAAAAATTGATATTACTTTGTTG 879
DB 5159 ATTGCACGAATCCTGTAACAATAGAAAGCGTTAGGAACAAGATTAAAGGAAAAATTTGGG 5218
QY 880 TATCGTGCATCTGGTCAAGAAATTCCTATTGGGGTGCACAGATTTTCCTCTAGTCCAGAA 939
DB 5219 AATAGTTCAGTTGCTCATCTTGAGATACCTTGATTTATATCAAGGTTTCATATGAATATTG 5278
QY 940 TATCATGAAGCCACAGTTAAATTCGTTAAAGTTTATAAAATCCACACCTTAAACACGGTGAT 999
DB 5279 ACTCATGAATCAAGGACGCTATTGCTAAGAATAAACAATATATACGACAAAAAGATATT 5338
QY 1000 ATCCATCATATGAATATTAAAGTTTTCAGCAACGGCTTAGATGATG 1045
DB 5339 TTGAACATTAAATGATTTTGATATTGACCGCTATATAACACTTGATG 5384

Query Match 3.2%; Score 35.2; DB 4; Length 1434;
Best Local Similarity 47.7%; Pred. No. 3.9;
Matches 103; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 774 AGCAGATAGTCTACCTGCCACACATATTAGAAGTGGTTCCAAATGACCATTTGAAAGCATTC 833
DB 552 AGCTTTTATTAGGTTTCTTCTCTAGAAAATAATACAAAACAAAGACATTTGTTCT 611
QY 834 TGAAGAAATCAGAAAAAGATAATGTTAAAAATTGATATTACTTTGTTGATCGTCATCTGG 893
DB 612 TGATAAAATTTCTGTAGTTATTCTACTTTTCGGCTTTGGTTAAATTTTATTCGCAITTAG 671
QY 894 TCNAGAAATTCATTTGGTGCACACAGATTTCTCTGTAGTCCAGATAATCATGAAGCCAC 953
DB 672 TAGCGTGAGTACGTTTGGTATCACCTCTCTCTGTCATTTGTGACATTTGTACTCGGTAT 731
QY 954 AGTTAAATTCGTTAAGTTTATAAAATCCACACCTTAA 989
DB 732 AGCCATTATCAATAATTTTACACGCGTCAGCTTAA 767

Query Match 3.2%; Score 35.2; DB 3; Length 6722;
Best Local Similarity 47.3%; Pred. No. 5.6;
Matches 107; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 820 ATTGAAGCATTCCTGAGAAATCAGAAAAAGATAATGTTAAAAATTGATATTACTTTGTTG 879
DB 5159 ATTGCACGAATCCTGTAACAATAGAAAGCGTTAGGAACAAGATTAAAGGAAAAATTTGGG 5218
QY 880 TATCGTGCATCTGGTCAAGAAATTCCTATTGGGGTGCACAGATTTTCCTCTAGTCCAGAA 939
DB 5219 AATAGTTCAGTTGCTCATCTTGAGATACCTTGATTTATATCAAGGTTTCATATGAATATTG 5278
QY 940 TATCATGAAGCCACAGTTAAATTCGTTAAAGTTTATAAAATCCACACCTTAAACACGGTGAT 999
DB 5279 ACTCATGAATCAAGGACGCTATTGCTAAGAATAAACAATATATACGACAAAAAGATATT 5338
QY 1000 ATCCATCATATGAATATTAAAGTTTTCAGCAACGGCTTAGATGATG 1045
DB 5339 TTGAACATTAAATGATTTTGATATTGACCGCTATATAACACTTGATG 5384

```
; LENGTH: 927
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-2636

Query Match      3.1%; Score 35; DB 4; Length 927;
Best Local Similarity 45.2%; Pred. No. 3.7;
Matches 128; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

QY  31 ATCATTCAGGTGACAAAGCTGTTGTTAAACAGATCTCTCAGTTCAGAAATTAAGGAG 90
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  529 ATGAATGAAAGAGCCCAAGCAATTAATGTACAAATATTTATGTAAAGACCAATGCTGAC 588
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  91 GGTACAGCCTTGGTGAAGGTTGAGGCTGTTGCTGTAACCACTGATTGGAAGCATATT 150
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  589 TCITTTGCCAATGATGGCATCTGAGCGGCTGCTGCATTTGTAATGATGATATATTTTA 648
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  151 GCTTATAAGATTTGGTCCAGAAAGTTCAATTTAGGATGTGACATTTGCTGTACAGTTGTC 210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  649 GCAGGCTTAATTTGCTAAGTCTTCTACACCTAAAGCAATTTGCAATTTGTTGCTCCAGTACTT 708
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  211 AACTTGGACCAATAGCTAGTACTGACTTGAAGTTTGGAGATACCGGTTTCGGTTTGT 270
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  709 TCATCAGAACCTTATGCAATTTATGATTGCAAAAGATGATCCAAATTTAAAGCAATTTGCT 768
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  271 CAGGTCCTTCCCAACAGATCCTTAAATAATGGTGCAATTTGCTG 313
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  769 GACGCTACTGTAATACTTATGGAATACTGGTCAAGTGGATG 811
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 19
US-09-328-352-3221
; Sequence 3221, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: GTC99-03PA
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 3221
; LENGTH: 4860
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-3221

Query Match      3.1%; Score 35; DB 4; Length 4860;
Best Local Similarity 57.9%; Pred. No. 7.2;
Matches 62; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY  150 TGCTTATAAGATTTGGTCCAGAAAGTTCAATTTCTAGGATGTGACATTTGCTGTGACAGTTGT 209
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  855 TGACGATAAAACTGCGCAATAGTTTAAACAGTTGCAATTTGATACAGCGGTAAGTAGT 914
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  210 CAACTTGGACCAATCTAGTACTGACTTGAAGTTGGAGATACCG 256
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  915 TAAAGGTGTAAGTGAAGCCCAATGCTGTTGTGACAGTTAAGAATGCTG 961
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 20
US-09-702-705-1408
; Sequence 1408, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
```

```
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702,705
; CURRENT FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 1833
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1408
; LENGTH: 360
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-702-705-1408

Query Match      3.1%; Score 34.8; DB 4; Length 360;
Best Local Similarity 51.3%; Pred. No. 2.8;
Matches 81; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY  927 TGCTAGTCCAGAAATATCATGAAGCCACAGTTAAATTCGTTAAAGTTTATAAATCCACACCT 986
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  200 TGGAGTCCAAATTTATCATGCTGCTGAGTGTAATAATATATGTTTCTTAAGACCGTGTGT 259
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  987 TAACACGGTGATATCCATCATATGATTAATAAGTTTTCAGCAACGGCTTAGATGATGT 1046
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  260 GAAAGAAATATAAATATCTCTTAAATAATGTTTATATTGAGTACATGTTGAAATAATTT 319
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  1047 CCCAGCTCTCACTGAAGGTATATAAAGAGGTAAAAACA 1084
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  320 TATATTGTGACACATTTGTTAAATAAATAATTTAAA 357
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 21
US-09-736-457-1408
; Sequence 1408, Application US/09736457
; Patent No. 6509448
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Aijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1408
; LENGTH: 360
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-736-457-1408

Query Match      3.1%; Score 34.8; DB 4; Length 360;
Best Local Similarity 51.3%; Pred. No. 2.8;
Matches 81; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY  927 TGCTAGTCCAGAAATATCATGAAGCCACAGTTAAATTCGTTAAAGTTTATAAATCCACACCT 986
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  200 TGGAGTCCAAATTTATCATGCTGCTGAGTGTAATAATATATGTTTCTTAAGACCGTGTGT 259
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  987 TAACACGGTGATATCCATCATATGATTAATAAGTTTTCAGCAACGGCTTAGATGATGT 1046
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  260 GAAAGAAATATAAATATCTCTTAAATAATGTTTATATTGAGTACATGTTGAAATAATTT 319
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  1047 CCCAGCTCTCACTGAAGGTATATAAAGAGGTAAAAACA 1084
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 320 TATATTGTGACACATTGTGTTAAATAAATAATATAAAA 357
RESULT 22
US-09-134-001C-461
; Sequence 461, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO. 461
; LENGTH: 1044
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-461
Query Match 3.1%; Score 34.8; DB 4; Length 1044;
Best Local Similarity 45.6%; Pred. No. 4.4;
Matches 123; Conservative 0; Mismatches 147; Indels 0; Gaps 0;
QY 499 CCGCAACATCTCATCTATTGATTTGGGTGGTCTACAGCAGTGGGTCAACACTA 558
Db 463 CCTAAGAAGGTGAACAGTAGTATTTCAGCAGCTTCAGTGCAGTTGGTTCAGTTGTG 522
QY 559 ATCCAAATGCCCAACATCAATGCTTATCTACTAAGATTGTAAGTGTCTCTTAAAG 618
Db 523 GGCCAAATCGAAGCTTAAAGGTTGCAGAGTCGTTGGTATAGCTGGTGAGATAAAAA 582
QY 619 CATGMAAAGCTTTTAAAGTCTTATGGTGGTGGTATGCTTTGACTATCATGATGCAGGC 678
Db 583 GTCAACTATCTAAAAATGAATCTCGTTTGTGCTGCTGATCGATTACAAAAAGATAAT 642
QY 679 GTTATTGAGCAGATCAATCGAAGTATCCAAACCTGCAACATGTTATTTGACGCTGTGGGA 738
Db 643 TTCCTGAAGCGTTAAAGAAGCGGTCCTAACCGTATAGATGCTACTTTCGAAATGTA 702
QY 739 AGCGAAGATAGTATCCCGAGGCGCTATAAA 768
Db 703 GGTGGATATATTGGCGATGAAGTCTTCAAA 732
RESULT 23
US-09-107-532A-3324/C
; Sequence 3324, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneka
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3324:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1053 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (8) LOCATION 1...1053
; SEQUENCE DESCRIPTION: SEQ ID NO: 3324:
US-09-107-532A-3324
Query Match 3.1%; Score 34.8; DB 4; Length 1053;
Best Local Similarity 52.0%; Pred. No. 4.4;
Matches 78; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
QY 920 GATTTCCTGCTAGTCCAGAAATATCATGAAGCCACAGTTAAATTCGTTAAGTTTATAAATC 979
Db 980 GTTTCACAGTCATCAAGACAGTCATGAACCAAGAACCTTCTTCTACCTGTTGATC 921
QY 980 CACACCTTAACACGGTGATATCCATCATATGATTAAGTTTTCAGCAACGGCTTAG 1039
Db 920 TTTCCATACACACAGTGTGACATCGATGTTATAGCGTGAATCGCTTGGGAATGACTGCG 861
QY 1040 ATGATGTCGCCAGCTCTCACTGAAGGTATAA 1069
Db 860 TCGTGGCATTTCTTCGCTAAAGTCAAA 831
RESULT 24
US/08/622
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1347 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 53..1168
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US/08/622,679D-11
Query Match 3.1%; Score 34.8; DB 3; Length 1347;
Best Local Similarity 48.5%; Pred. No. 4.9;
Matches 96; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
QY 680 TTATTGAGCAGATCAAAATCGAAGTATCCAAACCTGCAACATGTTATTGACGCTGTGGAA 739
Db 978 TTCTGGGCAAAATTCCTGCAGCAAGGAACCAATATTATTTTGGGATGATTTCATC 1037
QY 740 GCGAAGATAGTATCCCGAGGCGCTATAAAGTTCACAGCATAGTCTACCTGCCACATTAT 799

Db 1038 CCACTGAAGTAGTCAATGAAGCATATTCAAGATTAGCATATTCTCGCTTCTCAATTAC 1097
QY 800 TAGAAGTGGTTCACATGACCATTCAGAGCATTCCTGAAGAAATCAGAAAGATAATGTTA 859
Db 1098 TTGATGCTGATCTCTTGGCCATTGGCGCTTAACAGGCAAAAATGTCATGATAAAGTGA 1157
QY 860 AAATTGATATTACTTTGT 877
Db 1158 AGATAACAATAGACTGTAT 1175

RESULT 25
US-09-165-922A-11
; Sequence 11, Application US/09165922A
; Patent No. 6348641
; GENERAL INFORMATION:
; APPLICANT: STILES, JOHN I.
; MOISYADI, ISTEFO
; NEUPANE, KABI R.
; TITLE OF INVENTION: PURIFIED PROTEINS,
; RECOMBINANT DNA SEQUENCES AND PROCESSES FOR
; PRODUCING CAFFEINE FREE BEVERAGES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JONES, DAY, REAVIS & FOGUE
; STREET: NORTH POINT, 901 LAKESIDE AVENUE
; CITY: CLEVELAND
; STATE: OHIO
; COUNTRY: USA
; ZIP: 44114
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb

COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS-DOS V. 5.1
SOFTWARE: WordPerfect for Windows v. 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/165,922A
FILING DATE: 02-Oct-1998
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: GRIFFITH, CALVIN P.
REGISTRATION NUMBER: 34,831
REFERENCE/DOCKET NUMBER: 265036600003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (216) 586-7050
TELEFAX: (216) 579-0212
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1347 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna to mRNA
FEATURE:
NAME/KEY: CDS
LOCATION: 53..1168
SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-09-165-922A-11

Query Match 3.1%; Score 34.8; DB 4; Length 1347;
Best Local Similarity 48.5%; Pred. No. 4.9;
Matches 96; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
QY 680 TTATTGACGAGATCAAAATCGAAGTATCCAACTCGCAACATGTTATTGACGCTGTGGAA 739
Db 978 TTCTGGGCAAAATCCCTGCGAGCAACAGGAACCAATATTATTTTGGGATGATTCCATC 1037
QY 740 GCGNAGATGATATCCCGAGCCCTATAAAGTCACAGCATAGTCTACCTGCCACATTAT 799
Db 1038 CCACTGAAGTAGTCAATGAAGCATATTCAAGATTAGCATATTCTCGGTTATCTCAATTAC 1097
QY 800 TAGAAGTGGTTCACATGACCAATTGAAGCATTCCTGAAGAAATCAGAAAGATAATGTTA 859

Db 1098 TTGATGCTGATCTCTTGGCCATTGGCGCTTAACAGGCAAAAATGTCATGATAAAGTGA 1157
QY 860 AAATTGATATTACTTTGT 877
Db 1158 AGATAACAATAGACTGTAT 1175

RESULT 26
US-09-601-198-131/c
; Sequence 131, Application US/09601198
; Patent No. 6531583
; GENERAL INFORMATION:
; APPLICANT: Cassell, Gail H.
; APPLICANT: Chen, Ellison Y.
; APPLICANT: Glass, Jennifer S.
; APPLICANT: Glass, John I.
; APPLICANT: Heiner, Cheryl R.
; APPLICANT: Lefkowitz, Elliot
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREA PLASMA
; FILE REFERENCE: UAB-13452/22
; CURRENT APPLICATION NUMBER: US/09/601,198
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/073,189
; PRIOR FILING DATE: 1998-01-30
; NUMBER OF SEQ ID NOS: 181
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 131
; LENGTH: 1374
; TYPE: DNA
; ORGANISM: Ureaplasma urealyticum
US-09-601-198-131

Query Match 3.1%; Score 34.8; DB 4; Length 1374;
Best Local Similarity 43.3%; Pred. No. 4.9;
Matches 162; Conservative 0; Mismatches 212; Indels 0; Gaps 0;
QY 732 TGTGGGAAGCGAGATAGTATCCCGAGGCCTATAAAGTCACAGCATAGTCTACCTGC 791
Db 976 TGTGTGGATTATATATAAATAATCAAGTTTGTCTAACTTTAAGCATGGTATTTATGT 917
QY 792 CACATTATTAGAAAGTGGTTCCCAATGACCATTCGAAGCATTCCTGAAGAAATCAGAAAGA 851
Db 916 TAATTATGATGAATTAATAAATTAATGCGATTATTAATGTTTCATGAAGATTAAAGACCA 857
QY 852 TAATGTTAAATTTGATATTACTTTGTTGTTATCGTGCATCTGGTCAAGAAATTCATTGGG 911
Db 856 TTATATATCTAAAAATAACCCCTGTTGTTAATAATAATTTTGAATTTAAACCATCACAATAT 797
QY 912 TGCACAAGATTTCCTGCTAGTCCAGAAATATCATGAAGCCACAGTTAAATTCGTTAAGTT 971
Db 796 TAATAAGTAGTAGCATTAACGCATCTCGTAAATTTAGCAACATTAATATGGCTTTATTTT 737
QY 972 TATAATCCACACCTTTAACAACGGTGATATCCATCATATGAATATTTAAAGTTTTTCAGCA 1031
Db 736 AATGATTGTACACTACCAATGATTAAACAAGATAAAGCAATTAAGTTTATGATAA 677
QY 1032 CGGCTTTAGATGATGCCCGAGCTCTCACTGAAGGTATATAAAGAGGTATAAACAATAATGT 1091
Db 676 TGTATTGCTAAATTTGTTGATATTTCATGACGTTAAAAATTTTAGCATGATATGACAAAA 617
QY 1092 TAAGTATGTGCCA 1105
Db 616 TAATAACAATGCA 603

RESULT 27
US-09-173-053-3
; Sequence 3, Application US/09173053
; Patent No. 6451769
; GENERAL INFORMATION:
; APPLICANT: HUEBNER, Robert C.

```

; APPLICANT: NORMAN, Jon A.
; APPLICANT: LIANG, Xiaowu
; APPLICANT: LIANG, Kristin R.
; APPLICANT: CARNER, Alan G.
; APPLICANT: BARBOUR, Alan G.
; APPLICANT: LUKE, Catherine J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR ADMINISTERING BORRELIA DNA
; FILE REFERENCE: 454312-2440.1
; CURRENT APPLICATION NUMBER: US/09/173,053
; CURRENT FILING DATE: 1998-10-15
; PRIOR APPLICATION NUMBER: 08/563,998
; PRIOR FILING DATE: 1996-06-14
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 4016
; TYPE: DNA
; ORGANISM: Borrelia burgdorferi
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6)..(29)
; OTHER INFORMATION: N stands for A or G or C or T
; NAME/KEY: misc_feature
; LOCATION: (1078)..(1085)
; OTHER INFORMATION: N stands for A or G or C or T
US-09-173-053-3

```

RESULT 28
US-08-781-891-209
; Sequence 209, Application US/08781891
; Patent No. 6090620
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; APPLICANT: Yu, Chang-En
; APPLICANT: Oshima, Junko
; APPLICANT: Mulligan, John T.
; APPLICANT: Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
; TITLE OF INVENTION: WERNER'S SYNDROME
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; City: Seattle
; STATE: Washington

```

; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781.891
; FILING DATE: 27-DEC-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6090620century Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 240052.419
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 209:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 51259 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-781-891-209

Query Match 3.1%; Score 34.8; DB 3; Length 51259;
Best Local Similarity 49.2%; Pred. No. 21;
Matches 124; Conservative 0; Mismatches 122; Indels 6; Gaps 1;

Qy 187 TGTGACATTGCTGGTACAGTTGTCAAACCTTGGACCAAATGCTAGTACTGACTTGAAGGTT 246
Db 14400 TGGGATGTTATTGTTACAGTAGTCTGACATTTAACTAATCAGATTGTGTCATTTTGAAGTA 14459

Qy 247 GGAGATACCGGTTTCGGTTTGTTCACCGGTGTTCCCAACAGATCTCAAATAATGGTGCA 306
Db 14460 AATGTTACATTTTTTTTTTAAAGTAGTCCGGGTCTATAACAGAAAATAGCAAGCATCTTCA 14519

Qy 307 TTGTGCTGAATATGCCAGGGTTTATCCACTTTGTTTTTA-----CAAGAGTAACCTTAAC 360
Db 14520 TGGGGTGCCTTCCCAAGCGTACTTTGTGATTTGTCTTTTAACTTTTGGGAATAGACACTTGA 14579

Qy 361 CACTCAACTGCTGATGAATAATTTCTGAAGGCCCTGTGAAGAACTTCGAAATCTGCTGCAATCA 420
Db 14580 GGCAGATGCCTAATGAATCTCTACAGGACCTTGGAAAGCCCTTGAACCTTTTGCATTCA 14639

Qy 421 TTGCCAGTTTCG 432
Db 14640 GAGTGAATTTTG 14651

```

RESULT 29
US-09-618-166-209
Sequence 209, Application US/09618166
Patent No. 6583112
GENERAL INFORMATION:
APPLICANT: Fu, Ying-Hui
Yu, Chang-En
Oshima, Junko
Mulligan, John T.
Schellenberg, Gerald D.
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
WERNER'S SYNDROME
NUMBER OF SEQUENCES: 209
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed Intellectual Property Law Group
STREET: 701 Fifth Avenue, Suite 6300
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

c 89 34.4 3.1 418 11 US-09-991-936-213 Sequence 213, App
c 90 34.4 3.1 553 13 US-10-027-632-277990 Sequence 277990,

ALIGNMENTS

RESULT 1
US-10-081-644-1
; Sequence 1, Application US/10081644
; Publication No. US20020192782A1
; GENERAL INFORMATION:
; APPLICANT: Yamamoto, Hiroaki
; APPLICANT: Kimoto, No. US20020192782A1hiro
; TITLE OF INVENTION: NOVEL ENONE REDUCTASES, METHODS FOR
; TITLE OF INVENTION: PRODUCING SAME, AND METHODS FOR SELECTIVELY REDUCING A
; TITLE OF INVENTION: CARBON-CARBON DOUBLE BOND OF AN ALPHA, BETA-UNSATURATED KETONE
; TITLE OF INVENTION: USING THE REDUCTASES
; FILE REFERENCE: 06501-100001
; CURRENT APPLICATION NUMBER: US/10/081,644
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: JP 2001-49363
; PRIOR FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1113
; TYPE: DNA
; ORGANISM: Kluyveromyces lactis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1110)
US-10-081-644-1

Query Match 100.0%; Score 1113; DB 14; Length 1113;
Best Local Similarity 100.0%; Pred. No. 7.6e-284;
Matches 1113; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCTGAGTCCCAACCTCAAAAGCGCTCATCATTAAGAGTGCACAAAGCTGTGTAA 60
DB 1 ATCTGAGTCCCAACCTCAAAAGCGCTCATCATTAAGAGTGCACAAAGCTGTGTAA 60

QY 61 ACAGATGCTCAGTTCAGATTAAGAGGAGGACGCTTGGTGAAGGTTGAGGCTGTT 120
DB 61 ACAGATGCTCAGTTCAGATTAAGAGGAGGACGCTTGGTGAAGGTTGAGGCTGTT 120

QY 121 GCTGGTAACCAACTGATGGAAGCATATTGCTTATAAGATTGGTCCAGAAGGTTCAATT 180
DB 121 GCTGGTAACCAACTGATGGAAGCATATTGCTTATAAGATTGGTCCAGAAGGTTCAATT 180

QY 181 CTAGGATGACATGCTGCTGACATGCTCAAACTGGACCAATGCTAGTACTGACTTG 240
DB 181 CTAGGATGACATGCTGCTGACATGCTCAAACTGGACCAATGCTAGTACTGACTTG 240

QY 241 AAGTTGGAGATACCGGTTTCGGTTTGGTTCAGGCTCTCCCAACAGATCTCTAAAT 300
DB 241 AAGTTGGAGATACCGGTTTCGGTTTGGTTCAGGCTCTCCCAACAGATCTCTAAAT 300

QY 301 GGTGCAATTCCTGAATGACAGGTTTATCCACCTTTGTTTACAGAGTAATCTTAAT 360
DB 301 GGTGCAATTCCTGAATGACAGGTTTATCCACCTTTGTTTACAGAGTAATCTTAAT 360

QY 361 CACTCACTGCTGATGAATTTCTGAGGCCCTGTGAGAACTTCGAACTCTGCTGATCA 420
DB 361 CACTCACTGCTGATGAATTTCTGAGGCCCTGTGAGAACTTCGAACTCTGCTGATCA 420

QY 421 TTGCCAGTTTCGTTGACAACTGCTGTTGTTAGTTTGTGTCATCACTTGGGCTCAAAATG 480
DB 421 TTGCCAGTTTCGTTGACAACTGCTGTTGTTAGTTTGTGTCATCACTTGGGCTCAAAATG 480

QY 481 GAATGGCAACCACTACCCCGCAACATATCATCCATTAATTTGGGGTGTGCTACA 540
DB 481 GAATGGCAACCACTACCCCGCAACATATCATCCATTAATTTGGGGTGTGCTACA 540

QY 541 GCAGTGGTCAACAACCTAATCAAGTTGCCAACATATCAATGCTTATCTAAGATTGTA 600
DB 541 GCAGTGGTCAACAACCTAATCAAGTTGCCAACATATCAATGCTTATCTAAGATTGTA 600

QY 601 ACTGTTGCTTCTAAAAAGCATGAAAAGCTTTTAAAGCTTTATGGTCTGATGATGCTTT 660
DB 601 ACTGTTGCTTCTAAAAAGCATGAAAAGCTTTTAAAGCTTTATGGTCTGATGATGCTTT 660

QY 661 GACTATCATGATGCAGGCGTTATTGAGCAGATCAAAATCGAAGTATCCAAACCTGCAACAT 720
DB 661 GACTATCATGATGCAGGCGTTATTGAGCAGATCAAAATCGAAGTATCCAAACCTGCAACAT 720

QY 721 GTTATTGACGCTGTGGGAAGCAAGATAGTATCCCGAGCGCTTATAAAGTTCACAGCAGAT 780
DB 721 GTTATTGACGCTGTGGGAAGCAAGATAGTATCCCGAGCGCTTATAAAGTTCACAGCAGAT 780

QY 781 AGTCTACCTGCCACCATTTATTAGAAGTGTTCCTCAATGACCATTTCTCTGAAGAA 840
DB 781 AGTCTACCTGCCACCATTTATTAGAAGTGTTCCTCAATGACCATTTCTCTGAAGAA 840

QY 841 ATCAGAAAAGATAATGTTAAAAATTGATATCTTTGTTGATCGTGCATCTGGTCAAGAA 900
DB 841 ATCAGAAAAGATAATGTTAAAAATTGATATCTTTGTTGATCGTGCATCTGGTCAAGAA 900

QY 901 ATTCTATTGGGTGCAACAAGATTTCCTGCTAGTCCAGAAATATCATGAAGCCACAGTTAA 960
DB 901 ATTCTATTGGGTGCAACAAGATTTCCTGCTAGTCCAGAAATATCATGAAGCCACAGTTAA 960

QY 961 TTCGTTAAGTTTATAAATCCACACCTTAAACAAGGTGATATCCATCATATGATATAA 1020
DB 961 TTCGTTAAGTTTATAAATCCACACCTTAAACAAGGTGATATCCATCATATGATATAA 1020

QY 1021 GTTTTCAGCAACGGCTTAGATGATGCTCCAGCTCTCAGTGAAGGTATATAAAGAGTAA 1080
DB 1021 GTTTTCAGCAACGGCTTAGATGATGCTCCAGCTCTCAGTGAAGGTATATAAAGAGTAA 1080

QY 1081 AACAAAAATGTTAAAGTATGTTCCAGGTTTATA 1113
DB 1081 AACAAAAATGTTAAAGTATGTTCCAGGTTTATA 1113

RESULT 2
US-10-081-644-14
; Sequence 14, Application US/10081644
; Publication No. US20020192782A1
; GENERAL INFORMATION:
; APPLICANT: Yamamoto, Hiroaki
; APPLICANT: Kimoto, No. US20020192782A1hiro
; TITLE OF INVENTION: NOVEL ENONE REDUCTASES, METHODS FOR
; TITLE OF INVENTION: PRODUCING SAME, AND METHODS FOR SELECTIVELY REDUCING A
; TITLE OF INVENTION: CARBON-CARBON DOUBLE BOND OF AN ALPHA, BETA-UNSATURATED KETONE
; TITLE OF INVENTION: USING THE REDUCTASES
; FILE REFERENCE: 06501-100001
; CURRENT APPLICATION NUMBER: US/10/081,644
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: JP 2001-49363
; PRIOR FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 509
; TYPE: DNA
; ORGANISM: Kluyveromyces lactis
US-10-081-644-14

Query Match 45.1%; Score 502.2; DB 14; Length 509;
Best Local Similarity 98.4%; Pred. No. 1.9e-122;
Matches 501; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 160 ATTGGTCCAGAGGTTCAATTTAGGATGATGACATTTGGTACAGTGTGCAAACTTGA 219
DB 1 ATTGGTCCWAGGGVTCWATTTCTAGGATGATGACATTTGGTACAGTGTGCAAACTTGA 60

	Query Match	31.0%	Score 345.2	DB 14	Length 1122
	Best Local Similarity	58.5%	Pred. No. 1e-80		
	Matches 643	Conservative 0	Mismatches 448	Indels 9	Gaps 2
Qy	22	AAAGCCGTCAATCAATCAAGGTGACAAAGCTGTGTTTAAACAGATGCTCAGTTCACGAA	81		
Db	10	AAAGCTGTCTGTCATTGAAGACGGTAAAGCGGTGTCTAAAGAGGGCGTTCCTCATTCCTGAA	69		
Qy	82	TTAAAGAGAGGTTACAGCCTTGTGAAGGTTTGAGGCTGTTCCTGTAACCCAACTGATTGG	141		
Db	70	TTGGAAGAGGATTCTGATTATTGAATTAAGACACTCGCTGTCTGTAACCGACTGATTGG	129		
Qy	142	AAGCATATTCTTATTAAGATTTGGTCCAGAAGTTCAATTCTAGGATGTGACTGCTGGT	201		
Db	130	GCACACATTGACTATCAAGGTTCGGGCTCAAGATCTATTCTGGGATGTGACGTCGTCGGC	189		

;; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
;; FILE REFERENCE: AEWICA-X-2
;; CURRENT APPLICATION NUMBER: US/10/029,386
;; CURRENT FILING DATE: 2001-12-20
;; NUMBER OF SEQ ID NOS: 34288
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 25419
;; LENGTH: 572
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AL079336.5
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.62
;; OTHER INFORMATION: NT HIT: AF128626.1, EVALUATE 1.40e-01
US-10-029-386-25419

Query Match 3.7%; Score 41; DB 13; Length 572;
Best Local Similarity 45.7%; Pred. No. 2.1;
Matches 143; Conservative 0; Mismatches 170; Indels 0; Gaps 0;
QY 7 GTTCCACCCTCAAAAAGCGGTGATCATTTGAAGTGACAAAGCTGTTGTTAAACAGAT 66
DB 429 GTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 370
QY 67 GTCTCAGTTCCAGAAATTAAGAGGGTACAGCCTTGGTGAAGGTTGAGGCTGTTGCTGGT 126
DB 369 GATGTAGTGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 310
QY 127 AACCAACTGATGGAAGCATATGCTTATGAAGATTGGTCCAGAAAGGTTCAATTTCTAGGA 186
DB 309 GATGTAAGTGGTGTCAATATTGTTGATGATGATGATGATGATGATGATGATGATGATGAT 250
QY 187 TGTGACATGCTGGTACAGTTGTCAACTTGGACCAATGCTAGTACTGACTTGAAGTT 246
DB 249 GCTGATGCTGATGTTGGTGGTGTCAATATTGTTGATGATGATGATGATGATGATGATGAT 190
QY 247 GGAGATCCGGTTTCGGTTTTGTTTCACGGTGTCTCCCAACAGATCCTAAATAATGGTGCA 306
DB 189 GCTGATGTTGGTGGTGTCAATGTTGATGATGATGATGATGATGATGATGATGATGATGAT 130
QY 307 TTTGCTGAATG 319
DB 129 GATAGTGATGTTG 117

RESULT 8
US-10-312-841-1/c
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (3294164)
US-10-312-841-1

Query Match 3.7%; Score 41; DB 13; Length 3673778;
Best Local Similarity 47.5%; Pred. No. 3.2e+02;
Matches 122; Conservative 0; Mismatches 135; Indels 0; Gaps 0;
QY 857 TTAATAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 916

DB 3131933 TTAATAACAATTAACCGGTATAATATATTTTCTTTCTATATAAACACATAATATCATATAATC 3131874
QY 917 CAAGATTTCTGCTAGTCCAGAAATATCATGCAAGCCACAGTTAAATTCGTTAAGTTTATAA 976
DB 3131873 CATTTTATACTACTATAACAAATAATACAAACATTAATAACAAATAACAAATAATTTATTA 3131814
QY 977 ATCCACACCTTAAACAACGGTGATATCCATCATATGAATATTAAGTTTTCGCAACGGCT 1036
DB 3131813 CTATATATTTCTATAAATCAAAATCCAAATCAAAATCAATCAATCAATCAATCAATCAATCAAT 3131754
QY 1037 TAGATGATGTCGCCAGCTCTCAGTGAAGGTATAAAGAGGTAAAAACAATAATGTTAAGT 1096
DB 3131753 TTTATTACATCATCACACAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 3131694
QY 1097 ATGTTGCCAGGTTATAA 1113
DB 3131693 AACTTACCTTTTATAA 3131677

RESULT 9
US-10-312-841-2/c
; Sequence 2, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (379615)
US-10-312-841-2

Query Match 3.6%; Score 40.6; DB 13; Length 3673778;
Best Local Similarity 49.3%; Pred. No. 3.8e+02;
Matches 106; Conservative 0; Mismatches 109; Indels 0; Gaps 0;
QY 896 AAGAAATTTCTATGGGTGCAACAAAGATTTCCTGCTAGTCCAGAAATATCATGAGCCACAG 955
DB 694104 AAAAAATAAAATAAAATAAAATAATATACACAAAAACCAAAATATAATTTAAAAACAATTT 694045
QY 956 TTAATTCGTGATGTTTATAAATCCACACTTAACACCGGTGATATCCATCATATGAATA 1015
DB 694044 AAAAAATACTAAAAATAATATATCTTCAATAAATAAATAAATAAATAAATAAATAAATAA 693985
QY 1016 TTAAGTTTTCAGCAACGGCTTAGATGATGTCCACGCTCTCACTGAAGGTATATAAAGAAG 1075
DB 693984 ACATAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 693925
QY 1076 GTAAATAACAAATAATGTTAAGTATGTTGCCAGGTTA 1110
DB 693924 AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 693890

RESULT 10
US-09-864-761-20174/c
; Sequence 20174, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

Publication No. US20030143668A1
GENERAL INFORMATION:
APPLICANT: SUNA, MAKIKO
APPLICANT: ASAI, KIVOSHI
APPLICANT: AKIYAMA, YUTAKA
APPLICANT: ABURATANI, HIROYUKI
TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
FILE REFERENCE: 084335/0152
CURRENT FILING DATE: 2002-12-18
PRIOR FILING DATE: 2002-12-18
PRIOR APPLICATION NUMBER: JP 2001/246789
PRIOR FILING DATE: 2001-06-18
NUMBER OF SEQ ID NOS: 2430
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2179
LENGTH: 1168
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: source
LOCATION: (1)..(1168)
FEATURE:
NAME/KEY: CDS
LOCATION: (201)..(968)
US-10-017-161-2179

Query Match 3.4%; Score 38; DB 13; Length 1168;
Best Local Similarity 45.6%; Pred. No. 20;
Matches 134; Conservative 0; Mismatches 160; Indels 0; Gaps 0;
Qy 37 GAAGGTGACAAAGCTGTTTAAACACAGATGCTCAGTTCAGATTAAGAGGGGTACA 96
Db 284 GATGGTGATGATGATGCTGGTGATGATGCTGATGATGATGATGATGATGATGAT 343
Qy 97 GCCTTGGTGAAGGTGAGGCTGTTGCTGTAACCACTGATTCGAAGCATATTCCTTAT 156
Db 344 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 403
Qy 157 AAGATTGGTCAGAAAGGTTCAATCTAGGATGTGACATGCTGGTGACATGTTGTCAAACTT 216
Db 404 GATGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 463
Qy 217 GGACCAATGCTAGTACTGACTTGAAGGTTGGAGATACCGGTTTCGGTTTTTTCACGGT 276
Db 464 GGTGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 523
Qy 277 GCTTCCCAACAGATCCTAAAATGGTGCAATTTGCTGATATGCCAGGGTTTAT 330
Db 524 GATGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 577

RESULT 16
US-09-814-353-20029/c
Sequence 20029, Application US/09814353
Publication No. US20030165831A1
GENERAL INFORMATION:
APPLICANT: Lee, John
APPLICANT: Thompson, Pamela
APPLICANT: Lillie, James
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF OVARIAN CANCER
FILE REFERENCE: MRI-006B
CURRENT APPLICATION NUMBER: US/09/814,353
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: US 60/191,031
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: US 60/207,124
PRIOR FILING DATE: 2000-05-25
PRIOR APPLICATION NUMBER: US 60/211,940
PRIOR FILING DATE: 2000-06-15
PRIOR APPLICATION NUMBER: US 60/216,820
PRIOR FILING DATE: 2000-07-07

PRIOR APPLICATION NUMBER: US 60/220,661
PRIOR FILING DATE: 2000-07-25
PRIOR APPLICATION NUMBER: US 60/257,672
PRIOR FILING DATE: 2000-12-21
NUMBER OF SEQ ID NOS: 22037
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 20029
LENGTH: 10668
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: 10556, 10593, 10616, 10641, 10668
OTHER INFORMATION: n = A,T,C or G
US-09-814-353-20029

Query Match 3.4%; Score 37.6; DB 13; Length 10668;
Best Local Similarity 46.5%; Pred. No. 97;
Matches 121; Conservative 0; Mismatches 139; Indels 0; Gaps 0;
Qy 837 AGAAATCAGAAAGATAATGTTAAATTTGATATTACTTTTGTGTCGTCATCTGTCAC 896
Db 6607 AGTCAGCAGAGCAGCCACTGTTTGTCAAAATCTACTGTGTCCTTTAGCACCTTAAGCCCA 6548
Qy 897 AGAAATTTCTATTCGGTGCACAAAGATTTCTGCTAGTCCAGAAATATCATGAAGCCACAGT 956
Db 6547 GCTCTATCTATAAAGCAAGAAACAGTGACACACACTGCACATAATTTGTTAAGAAAGAAA 6488
Qy 957 TAAATTCGTTAAGTTTATAAATCCACACTTAACACGGTGATATCCATCATATGAATAT 1016
Db 6487 ACTATTTGCTTCCAGTTTAAATACCCAGCTATTACAAATGGAGAAATTTAAACACTGATCCT 6428
Qy 1017 TAAAGTTTTTCAGCAACGGCTTAGATGATGCCAGCTCTCACAGGTATATAAAGGAGG 1076
Db 6427 CAAAGTCTGCACCTTAACAAACTTTGACTTAAAGCCTGACTATAGTGTAAACCATCA 6368
Qy 1077 TAAAAACAAAAATGTTAAAGT 1096
Db 6367 TAAGACATAAATTAATAAT 6348

RESULT 17
US-09-790-988-1/c
Sequence 1, Application US/09790988
Patent No. US20020127687A1
GENERAL INFORMATION:
APPLICANT: SHIGENOBU, SHUJI
APPLICANT: WATANABE, HIDEMI
APPLICANT: HATTORI, MASAHIRA
APPLICANT: SAKAKI, YOSHIYUKI
TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
FILE REFERENCE: 081356/0159
CURRENT APPLICATION NUMBER: US/09/790,988
CURRENT FILING DATE: 2001-02-23
PRIOR APPLICATION NUMBER: JP2000-107160
PRIOR FILING DATE: 2000-04-07
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 640681
TYPE: DNA
ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match 3.4%; Score 37.6; DB 10; Length 640681;
Best Local Similarity 48.6%; Pred. No. 1e+03;
Matches 103; Conservative 0; Mismatches 109; Indels 0; Gaps 0;
Qy 716 AACATGTTATTCAGCGCTGCGGAGCGAAGATAGTATCCCGAGGCCCTATAAAGTCACAG 775
Db 547422 AATTAGTTATTGAGATGCGAACTAAACGGAATCTATTGATCAGAAAGAGCTATTTCGACGAG 547363
Qy 776 CAGATAGTCTACCTGCCACATATTAGAAAGTGGTTCCAATGACCATTTGAAAGCAATTCCTG 835

Db 547362 CAGCTACTATTTAGCAGAACAACTAGAACATTTGTTGATTTAAAGAGATATTCGTGCAAC 547303
Qy 836 AAGAAATCAGAAAGATATGTTAAATTTGATATATCTTTGTTGATCGTGCATCTGGTC 895
Db 547302 CTGAATTCAGAGAGAGAAACCTGAATTTGAACCTATTTGTTGATCGTCCAGTAGATGATT 547243
Qy 896 AAGAAATCTATTGGGTGCAACAAGATTTCCT 927
Db 547242 TAGAATTTACAGTTGCTGTTCTGCTAATTCCTT 547211

RESULT 18

US-10-029-386-25433
; Sequence 25433, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029.386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 25433
; LENGTH: 573
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL133232.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.16
; OTHER INFORMATION: NT HIT: U63807.1, EVALUATE 8.50e+00
US-10-029-386-25433

Query Match 3.4%; Score 37.4; DB 13; Length 573;
Best Local Similarity 45.4%; Pred. No. 19;
Matches 134; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

Qy 36 TGAAGGTGACAAAGCTGTTGTTAAACACAGATGTTCTCAGTTCACAGATTAAGGAGGTAC 95
Db 60 TGGTGGCGATTAATGTTGATGCAATGTTGGCGATGATGATGATAAGGTGATGCTGTC 119
Qy 96 AGCTTGGTGAAGTTGAGGCTGTTGTTGTAACCAACCTGATGTTGGAAGCATATGCTTA 155
Db 120 TGCTGCTGTTAATGGTAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 179
Qy 156 TAAGATTGGTCCAGAGGTTCAATTTCTAGGATGTGACATGCTGTTGATGATGATGATGAT 215
Db 180 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 239
Qy 216 TGGACCAAACTAGTACTGCTTCAAGGTTGGAGATACCGGTTTCGGTTTTCGTTTTCACGG 275
Db 240 TGATGATCATGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGA 299
Qy 276 TGCTTCCCAACACAGATCCTAAATATGTTGATGATGATGATGATGATGATGATGATGATGAT 330
Db 300 TGATGGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 354

RESULT 19

US-10-029-386-22786/c
; Sequence 22786, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2

; CURRENT APPLICATION NUMBER: US/10/029.386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 22786
; LENGTH: 835
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006549.28
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
; OTHER INFORMATION: SWISSPROT HIT: O29522, EVALUATE 1.80e+00
US-10-029-386-22786

Query Match 3.4%; Score 37.4; DB 13; Length 835;
Best Local Similarity 48.0%; Pred. No. 24;
Matches 107; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

Qy 107 AGTTGAGGCTGTTGCTGTAAACCAACTGATTTGGAAGCATATGCTTATAAGATGGTC 166
Db 568 AGCATGAGGATGCTGATCTCGGCTCGGCTGATGCTGATGATGATGATGATGATGATGATG 509
Qy 167 CAGAAAGTTCAATCTAGGATGTGACATGCTGCTGATGATGATGATGATGATGATGATG 226
Db 508 GTGATGCTGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 449
Qy 227 CTAGTACTGACTTTGAAGGTTGGAGATACCGGTTTCGGTTTTCGTTTTCACGGTGTCCCAA 286
Db 448 ATAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 389
Qy 287 CAGATCTAAATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 329
Db 388 CTGA 346

RESULT 20

US-10-369-493-42193
; Sequence 42193, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369.493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 42193
; LENGTH: 1200
; TYPE: DNA
; ORGANISM: Lactococcus lactis
US-10-369-493-42193

Query Match 3.3%; Score 37.2; DB 12; Length 1200;
Best Local Similarity 48.6%; Pred. No. 34;
Matches 102; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

Qy 900 AATTCTATGGGTGCAACAGATTTCTGCTAGTCCAGATATCATGACCCACAGTTAA 959
Db 165 AATTACATTTGGAACACAGGACTAGTCTGCTTGAGAGAGATAAAAAAGTCAACTTTA 224
Qy 960 ATTCTGTTAAGTTTATAAATCCACACCTTAACACCGGTGATATCCATCATATGAATATTA 1019

Db 225 CTCGAAAAATTTTCCTAAATTAGCGTCATTTGAATTTGATTGATGAAGTTGAAAAAGAA 284
Qy 1020 AGTTTTCAGCAACGGCTTAGATGATGCCAGCTCTCAGTGAAGGTATAAAGAGGTAA 1079
Db 285 AGATGGTGAGCTTTGGTCTAATATGTTAAAGGATGATGTCATGCTTAAAGGAGCAGG 344
Qy 1080 AACAACAAAATGTTAAGTATGTTGCCAGGTT 1109
Db 345 ATATGAATTCACAAAGGTTTGAATGTT 374

RESULT 21

US-10-252-157-159/c
; Sequence 159, Application US/10252157
; Publication No. US20030190640A1
; GENERAL INFORMATION:
; APPLICANT: Paris, Mary
; APPLICANT: Pearson, Cecelia I.
; TITLE OF INVENTION: GENES EXPRESSED IN PROSTATE CANCER
; FILE REFERENCE: PA-0027-1 US
; CURRENT APPLICATION NUMBER: US/10/252.157
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: 60/295,048
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 501
; SOFTWARE: PERL Program
; SEQ ID NO 159
; LENGTH: 7536
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030190640A1 903479.3
; LOCATION: 3017-3175, 5473, 5489, 5500, 5504, 6570
; OTHER INFORMATION: a, t, c, g, or other
US-10-252-157-159

Query Match 3.3%; Score 37.2; DB 13; Length 7536;
Best Local Similarity 51.9%; Pred. No. 1e+02;
Matches 84; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
Qy 936 AGAATATCATGAAGCCACAGTAAATTCGTTAAAGTTTATAATCCACACCTTAAACAAGG 995
Db 2834 ATATTCCTGATATGCAATTTAAAAAATAAAGTTATGAGACAAATATGAATTAAGT 2775
Qy 996 TGATATCATCATATGAATATAAAGTTTTCAGCAACGGCTTAGATGATGTCACAGCTCT 1055
Db 2774 CAATTTACAGTTATAAATATTTCAATTTACAGCTACACACTATAAATTAACAGGAATAA 2715
Qy 1056 CACTGAAGGTATAAAGAGGTAAACAAACAAATGTTAAGTA 1097
Db 2714 AATTCAGTACAGAAAGAGGAGGGGAAAAATAAAGCA 2673

RESULT 22

US-10-198-846-14020/c
; Sequence 14020, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198.846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18

; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14020
; LENGTH: 7772
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-198-846-14020

Query Match 3.3%; Score 37.2; DB 15; Length 7772;
Best Local Similarity 51.9%; Pred. No. 1e+02;
Matches 84; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
Qy 936 AGAATATCATGAAGCCACAGTAAATTCGTTAAAGTTTATAAATCCACACCTTAAACAAGG 995
Db 3446 ATATTCCTGATATGCAATTTAAAAAATAAAGTTATGAGACAAATATGAATTAAGT 3387
Qy 996 TGATATCATCATATGAATATAAAGTTTTCAGCAACGGCTTAGATGATGTCACAGCTCT 1055
Db 3386 CAATTTACAGTTATAAATATTTCAATTTACAGCTACACACTATAAATTAACAGGAATAA 3327
Qy 1056 CACTGAAGGTATAAAGAGGTAAACAAACAAATGTTAAGTA 1097
Db 3326 AATTCAGTAAACAGAAAGAGGAGGGAAAAATAAAGCA 3285

RESULT 23

US-09-788-711A-3/c
; Sequence 3, Application US/09788711A
; Patent No. US20020058328A1
; GENERAL INFORMATION:
; APPLICANT: Tania Tamsin Testa
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP-30225
; CURRENT APPLICATION NUMBER: US/09/788.711A
; CURRENT FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 0004196.2
; PRIOR FILING DATE: 2000-02-19
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 8772
; TYPE: DNA
; ORGANISM: HOMO SAPIENS
US-09-788-711A-3

Query Match 3.3%; Score 37; DB 9; Length 8772;
Best Local Similarity 58.7%; Pred. No. 1.2e+02;
Matches 64; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
Qy 11 CAACCACTCAAAAAGCCGTCATCTTGAAGGTGACAAAGCTGTTGTTAAAAACAGATGCT 70
Db 5456 CAGCTCGAGGATAGCTGTGCCAGTCTGTTGCTGCAATAGCTGTAGCAGACACAGGTTT 5397
Qy 71 CAGTTCAGAAATTAAGAGGGGTACAGCCTTGCTGAAGTTGAGGCTGT 119
Db 5396 GAGTCACAAGGGTCAGGAGGTACAGCCTTGCTCCACGTTGATGCTCT 5348

RESULT 24

US-09-788-711A-1/c
; Sequence 1, Application US/09788711A
; Patent No. US20020058328A1
; GENERAL INFORMATION:
; APPLICANT: Tania Tamsin Testa
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP-30225
; CURRENT APPLICATION NUMBER: US/09/788.711A
; CURRENT FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 0004196.2
; PRIOR FILING DATE: 2000-02-19
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1

[illegible]

```

RESULT 30
US-09-815-242-6606
; Sequence 6606, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haeelbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIORITY APPLICATION NUMBER: 60/191,078
; PRIORITY FILING DATE: 2000-03-21
; PRIORITY APPLICATION NUMBER: 60/206,848
; PRIORITY FILING DATE: 2000-05-23
; PRIORITY APPLICATION NUMBER: 60/207,727
; PRIORITY FILING DATE: 2000-05-26
; PRIORITY APPLICATION NUMBER: 60/242,578
; PRIORITY FILING DATE: 2000-10-23
; PRIORITY APPLICATION NUMBER: 60/253,625
; PRIORITY FILING DATE: 2000-11-27
; PRIORITY APPLICATION NUMBER: 60/257,931
; PRIORITY FILING DATE: 2000-12-22
; PRIORITY APPLICATION NUMBER: 60/269,308
; PRIORITY FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6606
; LENGTH: 969
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1) ... (969)

```

US-09-815-242-6606

Query Match	3.3%	Score 36.6;	DB 9;	Length 969;
Best Local Similarity	58.9%	Pred. No. 43;		
Matches 63;	Conservative 0;	Mismatches 44;	Indels 0;	Gaps 0;

Qy	37	GAAGGTGACAAAGCTGTTGTTAAACACAGATGCTCAGTTCAGATTCCAGATTAAAGGAGGTACA	96
Db	55	GGAAAGTAAAGAGGTACTTTGAAGAAGCAGAGTAACACTGCCAGATTATCGGAACATCAA	114
Qy	97	GCCTTGGTGAAGGTTGAGGCTGTTGCTGGTAACCCCACTGATTGGAA	143
Db	115	GTGTTGGTAAAGAATACCGACATCAATTAAATCCGATTGATTGGAA	161

Search completed: January 10, 2004, 10:03:48
Job time : 477 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2004, 06:18:49 ; Search time 2773 Seconds
(without alignments)

9755.096 Million cell updates/sec

Title: US-10-081-644-1

Perfect score: 1113

Sequence: 1 atgcagttccaccactca.....agatgttcaggattataa 1113

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database :

EST:

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpi:*
7: em_estro:*
8: em_estl:*
9: gb_est1:*
10: gb_est2:*
11: gb_estc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	294.4	26.5	1011	29	CNS06N13
C 2	182.4	16.4	896	29	CNS06LDW
C 3	181.2	16.3	881	29	CNS06LDV
C 4	154.6	13.9	1072	29	CNS06LTC

C 5	146.6	13.2	488	28	AZ931755
C 6	137.2	12.3	1118	29	CNS06LW5
C 7	132.6	11.9	434	28	AZ931754
C 8	126.6	11.4	454	28	AQ501781
C 9	125.6	11.3	546	28	AQ873738
C 10	125	11.2	537	28	AQ873737
C 11	124	11.1	529	28	AQ502797
C 12	123.8	11.1	467	29	BZ300244
C 13	119	10.7	946	28	CNS06HG8
C 14	117.8	10.6	542	28	AQ873736
C 15	115.4	10.4	337	14	T38505
C 16	108.4	9.7	540	28	AZ926593
C 17	97.6	8.8	404	28	AZ926284
C 18	96.4	8.7	353	28	AQ501976
C 19	90.2	8.1	539	28	AQ876344
C 20	79.6	7.2	522	28	AZ910325
C 21	73.2	6.6	574	29	BZ293919
C 22	61.2	5.5	454	28	AQ501614
C 23	56.2	5.0	1146	29	CNS07BB3
C 24	52.2	4.7	556	13	BQ143479
C 25	51.6	4.6	515	12	B1516433
C 26	51.6	4.6	595	12	BP099113
C 27	51.6	4.6	619	12	B1505803
C 28	51.6	4.6	657	12	B1503620
C 29	51.6	4.6	669	12	B1504612
C 30	51.6	4.6	706	12	B1508141
C 31	48.6	4.4	1035	13	BX335830
C 32	47.2	4.2	542	10	BE920514
C 33	47.2	4.2	620	10	BQ965222
C 34	46.6	4.2	541	13	BQ090513
C 35	46.6	4.2	544	12	B1480612
C 36	46.6	4.2	925	28	BH155926
C 37	46.2	4.2	1101	29	CNS00FVE
C 38	45.8	4.1	315	14	C90538
C 39	45.8	4.1	389	10	BE923966
C 40	44.8	4.0	514	9	AJ273783
C 41	44.8	4.0	532	29	BZ289933
C 42	44.8	4.0	775	13	BQ517113
C 43	44.8	4.0	572	14	CB014235
C 44	44.6	4.0	753	29	BZ866896
C 45	44.6	4.0	416	10	BZ324284
C 46	44.2	4.0	463	10	BE923969
C 47	44.2	4.0	624	10	BG521829
C 48	44.2	4.0	694	28	AZ522107
C 49	44.2	4.0	554	28	BZ322459
C 50	44	4.0	533	28	BH872432
C 51	43.4	3.9	538	29	BZ902429
C 52	43	3.9	554	9	AW648111
C 53	43	3.9	584	14	CB014069
C 54	43	3.9	597	14	CB014844
C 55	43	3.9	602	14	CB013742
C 56	43	3.9	675	29	BZ833475
C 57	43	3.9	706	29	BZ842180
C 58	43	3.9	711	29	BZ881273
C 59	43	3.9	809	29	BZ843883
C 60	43	3.9	840	10	BZ922932
C 61	42.6	3.8	440	10	BH184140
C 62	42.6	3.8	482	28	CNS07F5F
C 63	42.6	3.8	962	13	BU522130
C 64	42.6	3.8	1101	29	CNS0039G
C 65	42.6	3.8	529	10	BE462251
C 66	42.4	3.8	537	10	BG642450
C 67	42.4	3.8	578	12	B1929612
C 68	42.4	3.8	594	10	BF097226
C 69	42.4	3.8	631	29	BZ951629
C 70	42.4	3.8	668	9	AI775773
C 71	42.4	3.8	683	9	AW624103
C 72	42.4	3.8	700	12	B1927387
C 73	42.4	3.8	784	12	B1931636
C 74	42.4	3.8	798	29	BZ838591
C 75	42.4	3.8	823	29	BZ839738
C 76	42.4	3.8	879	28	BH151807
C 77	42.4	3.8			

AZ931755	474	dhz89
AL404715	T3 end of	
AZ931754	474	dhz89
AQ501781	V14B11	mt
AQ873738	V74E3	mt
AQ873737	V75E1	mt
AQ502797	V75D10	mt
BZ300244	KD0613	R1
AL398958	T7 end of	
AQ873736	V74E1	mt
T38505	EST104002	S
AZ926593	476	d1809
AZ926284	476	d1808
AQ501976	V11A3	mt
AQ876344	V98D9	mt
AZ930225	474	dhz51
BZ293919	CG0601	r1
AQ501614	V17P2	mt
AL437653	T3 end of	
BQ143479	fmh1c	pk0
B1516433	BE160022A	
BP099113	BP099113	
B1505803	BB170024A	
B1503620	BB170018A	
B1506164	BB170014B	
B1504612	BB170025B	
B1508141	BB170011A	
BX335830	BX335830	
BE920514	EST424283	
BQ965222	EST461041	
BQ090513	rc66d10	Y
B1480612	EST0065	H
BH155926	ENTTJ37TF	
AL071298	Drosophila	
C90538	C90538	Dicc
BE923966	EST427735	
AJ273783	AJ273783	
BZ289933	CH240	32L
BQ517113	EST624528	
CB014235	As tgz	76
BZ866896	CH240	I91
BE924284	EST428053	
BE923969	EST427738	
BG521829	15-14	Ste
AZ522107	202PBC01	
BZ922459	CH240	121
BH872432	hp0562	b
BZ902429	CH240	22P
AW648111	EST326565	
CB014069	As tgz	73
CB014844	As tgz	86
CB013742	As tgz	69
BZ833475	CH240	270
BZ842180	CH240	213
BZ881273	CH240	279
BZ843883	CH240	227
BE922932	EST426701	
BH184140	024	J
AL621093	T7 end of	
BU522130	AGENCOURT	
AL063921	Drosophila	
BE462251	EST324515	
BG642450	EST355926	
B1929612	EST549501	
BF097226	EST400604	
BZ951629	CH240	44A
AI775773	EST256873	
AW624103	EST322048	
B1927387	EST547276	
B1931636	EST551525	
BZ838591	CH240	225
BZ839738	CH240	192
BH151807	ENTQ16TF	

```

78 42.4 3.8 1048 13 BX354430
c 79 42.2 3.8 994 13 BX414650
c 80 42 3.8 556 29 BZ869915
c 81 42 3.8 1101 29 CNS00001
c 82 41.8 3.8 326 14 C92181
c 83 41.8 3.8 1201 13 BX384819
c 84 41.6 3.7 308 9 A1374249
c 85 41.6 3.7 465 29 BZ955772
c 86 41.6 3.7 617 12 BJ371293
c 87 41.6 3.7 712 13 BQ94346
c 88 41.4 3.7 238 29 BZ927574
c 89 41.4 3.7 425 29 BZ913257
c 90 41.4 3.7 468 29 BZ914511

```

ALIGNMENTS

```

RESULT 1
CNS06N13/c 1011 bp DNA linear GSS 17-JUN-2001
LOCUS T7 end of clone AU0AA008F10 of library AU0AA from strain CBS 3082
DEFINITION of Saccharomyces kluyveri, genomic survey sequence.
ACCESSION AL406189
VERSION AL406189.1 GI:12169887
KEYWORDS GSS.
SOURCE Saccharomyces kluyveri
ORGANISM Saccharomyces kluyveri

```

```

REFERENCE
AUTHORS Souciet J.L., Aigle M., Artiguenave F., Blandin G.,
Bolotin-Fukuhara M., Bon E., Brottier P., Casaregola S.,
de-Montigny J., Dujon B., Durrens P., Lepingle A., Llorente B.,
Malpertuy A., Neuveglise C., Ozier-Kalogeropoulos O., Potier S.,
Saurin W., Tekala F., Toffano-Nioche C., Wesolowski-Louvel M.,
Wincker P. and Weissenbach J.

```

```

TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 1152876
PUBMED 20584711

```

```

REFERENCE
AUTHORS Neuveglise C., Bon E., Lepingle A., Wincker P., Artiguenave F.,
Gallardin C. and Casaregola S.
TITLE Genomic exploration of the hemiascomycetous yeasts: 9.
JOURNAL Saccharomyces kluyveri
MEDLINE 20584719
PUBMED 1152884

```

```

REFERENCE
AUTHORS 3 (bases 1 to 1011)
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. [E-mail :
sereq@genoscope.cns.fr - Web : www.genoscope.cns.fr]
COMMENT This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
Location/Qualifiers
1..1011
/organism="Saccharomyces kluyveri"
/mol_type="genomic DNA"
/strain="CBS 3082"
/db_xref="taxon:4934"
/clone="AU0AA008F10"

```

```

misc_feature /clone lib="AU0AA"
/notes="end : T7"
complement (<185..>1007)
/notes="similar to Saccharomyces cerevisiae ORF YNL134c (
similarity to C.carbonum tox d gene )
1 putative frameshift(s)"
/evidence=not experimental
misc_feature complement (<188..>1007)
/notes="similar to Saccharomyces cerevisiae ORF YCR102c (
similarity to C.carbonum tox d gene )
1 putative frameshift(s)
similar to Saccharomyces cerevisiae ORF YLR460c (
similarity to C.carbonum tox d protein )"
/evidence=not experimental
BASE COUNT 271 a 222 c 215 g 295 t 8 others
ORIGIN

```

```

Query Match 26.5%; Score 294.4; DB 29; Length 1011;
Best Local Similarity 61.3%; Pred. No. 2.1e-62;
Matches 504; Conservative 2; Mismatches 314; Indels 2; Gaps 2;
Qy 292 CCTAAAAATGGTGCATTTGCTGAATATGCCAGGGTTTATCCACCTTTGTTTACAGAGT 351
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1004 CCTGAGAATGGTGGCTTTGCTGAATACAGCCGATTGGTCTCTTCTTATTAAGGCA 945
Qy 352 AACTTAACCTCACTCAACTGCTGATGAATTTCTGAAGGCCCTGTGAAGACTTCGATCT 411
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
944 GAGTTAACACCTAGTACMAAGRAT-ACATACAGAAAGGTGTGGTCAAAACCTTTGAGTCG 886
Qy 412 GCTGCATCATTTGCCAGTTCCTGTTGACAACTGCTGGTGTAGTTGTGTCATCATCTGGGC 471
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
885 GACTB-TCGCTACTCTGCTTGCCTTGACACCGGCTGGTGTGCTTAAAGTCACTACCTGGGC 827
Qy 472 TCAAAATGGAATGGACCCATCTACCCGCAACATCACTCATCCATATTTGATTTGGGGT 531
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
826 AGCAAGTTTGAATGGGAACCATCAAGCCCTCAACATGATTTCCCTCTCCTGATTTGGGGT 767
Qy 532 GGTGCTACAGAGTGGTCAACCACTAATCCAGTTGCCAACAATCATCATGCTTATCT 591
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
766 GGTGCTACAGTGTGGTCAATTTTGTGTTGCTTAAAGTCACTTATTAAGTCTTATGCTGAT 707
Qy 592 AAGATTGTAACTGCTCTCTTAAAGAGCATGAAAGCTTTTAAAGTCTTATGCTGCTGAT 651
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
706 AAGATTGTGGCGTCTCTTCCAGAAACACGATCCATCTTGAAGCAATACGCTGCTGAT 647
Qy 652 GATGCTTTGACATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 711
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
646 GAGATCTTCGATTTACCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 597
Qy 712 CTGCAACATGTTATTCACGCTGTGGGAAGCAAGATAGTATCCCGAGGCTTATTAAGTC 771
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
586 CTACAACTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 527
Qy 772 ACAGCAGATAGTCTACTCTGCCCATTTATTAGAGTGGTTCCCAATGACCATTTGAAGCAT 831
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
526 ACTCCAGATAACTAGCAACACCATTTGTCGATTTGGCTGCTTTAAGAGGGGGACAT 467
Qy 832 CTGGAAGAAATCAGAAAAGATATGTTAAATGATATTAATTTGTTGTTGATGATGAT 891
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
466 GAAAAGAGAAAGAAAGAAATCAATGTTAAGATGGAGCAGCCCTTGTATATCTTGGCTGCC 407
Qy 892 GGTCAAGAAATCTATTGGTGCACCAAGATTTCTGCTAGTCCAGCAATATCATGAAGCC 951
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
406 GGCTAGAAAGTGCATTTCCGCCCCCACTATCTCTGCTAACCACCAAGATACAGAGGGCG 347
Qy 952 ACAGTTAAATTCGTTAAGTTTAAATCCACACTTTAAACACCGGTGATATCCATCATATG 1011
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
346 ACTAGAAGATTCGTGGAATTTGTCAACCCCAAGATCAACACCGTGAATTCACCATCT 287
Qy 1012 AATATTAAAGTTTTCAGCAACCGCTTAGATGATGATGATGATGATGATGATGATGAT 1071
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
286 CTAATCCGTGTTTTTGAACACGGTTGGAAGATGTTCTCTGGACTTACTGCTGAATCCAA 227

```

```

Qy 1072 GAAGGTAACAAAATGTTAAGTATGTTCCAGGTTATAA 1113
Db 226 CGAGGTAAGATTCGGGTGTCARATTTGTCAGCTTGAA 185

RESULT 2
CNS06LDW/c
LOCUS
DEFINITION
T3 end of clone AT0AA012G10 of library AT0AA from strain CBS 4311
of Saccharomyces servazzii, genomic survey sequence.
ACCESSION
AL404058
VERSION
AL404058.1 GI:12165095
KEYWORDS
GSS.
SOURCE
Saccharomyces servazzii
ORGANISM
Saccharomyces servazzii
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE
1 (bases 1 to 896)
Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Lorente,B.,
Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)
20584711
11152876
2 (bases 1 to 896)
Casaregola,S., Lepingle,A., Bon,E., Neuvéglise,C., Nguyen,H.,
Artiguenave,F., Wincker,P. and Gaillardin,C.
Genomic exploration of the hemiascomycetous yeasts: 7.
Saccharomyces servazzii
FEBS Lett. 487 (1), 47-51 (2000)
20584717
11152882
3 (bases 1 to 896)
Genoscope.
Direct Submission
Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
segre@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
Location/Qualifiers
1..896
/organism="Saccharomyces servazzii"
/mol_type="genomic DNA"
/strain="CBS 4311"
/db_xref="taxon:27293"
/clone="AT0AA012G10"
/clone_lib="AT0AA"
/note="end : T3"
complement(<282..>892)
/note="similar to Saccharomyces cerevisiae ORF YNL134c [
similarity to C.carbonum toxD gene ]
1 putative frameshift(s)"
/evidence=not experimental
misc_feature
252 a. 136 c 190 g 316 t 2 others
BASE COUNT
ORIGIN

Query Match 16.4%; Score 182.4; DB 29; Length 896;
Best Local Similarity 56.9%; Pred. No. 1.7e-34;
Matches 350; Conservative 1; Mismatches 263; Indels 1; Gaps 1;

RESULT 3
CNS06LDV
LOCUS
DEFINITION
T7 end of clone AT0AA012G10 of library AT0AA from strain CBS 4311
of Saccharomyces servazzii, genomic survey sequence.
ACCESSION
AL404057
VERSION
AL404057.1 GI:12165093
KEYWORDS
GSS.
SOURCE
Saccharomyces servazzii
ORGANISM
Saccharomyces servazzii
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE
1 (bases 1 to 881)
Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Lorente,B.,
Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)
20584711
11152876
2 (bases 1 to 881)
Casaregola,S., Lepingle,A., Bon,E., Neuvéglise,C., Nguyen,H.,
Artiguenave,F., Wincker,P. and Gaillardin,C.

```

```

TITLE      Genomic exploration of the hemiascomycetous yeasts: 7.
JOURNAL    Saccharomyces servazzii
MEDLINE    FEBS Lett. 487 (1), 47-51 (2000)
PUBMED     20584717
REFERENCE  3 (bases 1 to 881)
AUTHORS    Direct Submission
TITLE      Direct Submission
JOURNAL    Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
COMMENT     2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
            This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES   Location/Qualifiers
            source
              1..881
              /organism="Saccharomyces servazzii"
              /mol_type="genomic DNA"
              /strain="CBS 4311"
              /db_xref="taxon:27293"
              /clone="ATOAA012G10"
              /clone_lib="ATOAA"
              /note="end : T7"
            misc_feature
              <140..>877
              /note="similar to Saccharomyces cerevisiae ORF YLR460c [
              similarity to C.carbonum toxD protein ]"
              /evidence=not_experimental
            misc_feature
              <143..>877
              /note="similar to Saccharomyces cerevisiae ORF YNL134c [
              similarity to C.carbonum toxD gene ]"
              /evidence=not_experimental
            misc_feature
              <158..>877
              /note="similar to Saccharomyces cerevisiae ORF YCR102c [
              similarity to C.carbonum toxD gene ]"
              /evidence=not_experimental
BASE COUNT 236 a 231 c 205 g 209 t
ORIGIN
Query Match      16.3%; Score 181.2; DB 29; Length 881;
Best Local Similarity 54.7%; Pred. No. 3.3e-34;
Matches 404; Conservative 0; Mismatches 328; Indels 6; Gaps 2;
Qy      1  ATGTCAGTTCCCAACCACTCAAAAGCGCTCATCTATTCGAAGTGACAAAGCTGTGTTTAA 60
Db      1  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      61  ACAGATGCTCTCAGTTCAGTAATTAAGGAGGTTACAGCTTGTGTAAGGTTGAGGCTGTT 120
Db      61  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      200  AC---CGTCCATTGCTCACTGACGACGAATACGCTTGTATCAGACCAAGCGGTT 256
Db      200  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      121  GCTGGTAACCAACTGATGGAACATATTGCTTATTAAGATTGCTCAGAAAGGTTCAATT 180
Db      121  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      257  GCCGGTAACCAACCGATTGGAACATGTCAAGTTCGGCCCTAGGTTCCAAACGGGCGGTC 316
Db      257  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      181  CTAGAGTGACATGCTGCTGTTGACAGTTGTCACAACTTCGACCAATGCTAGTCTGAC--- 237
Db      181  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      317  ATCGGTGATAGTCGCCCGGTTGAATCGTCAAGTTAGGTTCCAAAGTCGACAAACAAA 376
Db      317  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      238  TTGAAGGTTGAGATACCGGTTTCGGTTTGTTCACCGTGTCTCCAAACAGATCTCTAAA 297
Db      238  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      377  TTCCACGTAGTGACTATGCTACGGGTTGCTTCATGGGCTCTCTAGACAGCTCCAGAT 436
Db      377  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      298  AATGGTGCAATTTGCTGAATATGCCAGGGTTATCCACCTTTGTTTTCAGAGTAACCTTA 357
Db      298  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      437  AATGGGGCTTTGAGAATAGCTGCTAGATTCCATGATCACCCTTCCCTGTCCCCCAAC 496
Db      437  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      358  ACTCACTCAACTGCTGATGAAATTTCTGAAGGCCCTGTGAAGAACTTCGAATCTGCTGCA 417
Db      358  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      497  GGTCTTTCCGGCAAGGATTCATTCAGAGGGCCCGTGACTCTCTTGAAGCGGTGCC 556
Db      497  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      418  TCATTGCCAGTTTTCGTGACAACTGCTGGTGTGTTAGTTTGTGTCATCATCTTTGGGCTCAAAA 477
Db      418  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      557  ACGATTCCATGTTCTCTGTTGACTGTCAGGCGCCACTCTATTCCACCACATGGGTCTCAAC 616
Db      557  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      478  ATGGAATGGCAACCATCTACCCCGCAACATACCTATCCATTATGTTTGGGTTGGTGTCT 537
Db      478  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      617  ATGGAGTGGGAACCAAGAGGTCCTCAAGAAAGGGAAGTGTGTTTGTATCTGGGGGGTGTCT 676
Db      617  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      538  ACAGCAGTGGGTCAACAACTAAATCCCAAGTTGCCAAACATATCAATGCTTATATCAAGATT 597
Db      538  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      677  ACGCTCTGTCCTTCCACTTTTACAACTATTGCCAAAGTTAATGCTTCCGAGAGATC 736
Db      677  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      598  GTAACCTGTTGCTTCTAAAAAGCATGAAAGCTTTTAAAGTCTTTATGGTGTCTGATGATGC 657
Db      598  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      737  ATTGTCGTGGCTCCAAAGAAATGAAGCTAAATTAAGAAATACGGTGTGCGATGATGTTA 796
Db      737  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      658  TTTGACTATCATGCTCAGCGGTTATGACAGATCAATTCGAAGTATCCAAACCTGCAC 717
Db      658  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      797  TTCGACTATCAGATCCGACGTTATTGACCAATCAAGGCCAATACACCGATTGGTT 856
Db      797  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      718  CATGTTATTGACGCTGTG 735
Db      718  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      857  TGGTTATTAGATTGTTG 874
Db      857  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT 4
LOCUS     CNS06LTC
DEFINITION 1072 bp DNA linear GSS 17-JUN-2001
ACCESSION AL404614
VERSION    AL404614.1 GI:12166302
KEYWORDS   GSS.
SOURCE     Saccharomyces servazzii
ORGANISM   Saccharomyces servazzii
REFERENCE  1 (bases 1 to 1072)
AUTHORS    Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissensbach,J.
            Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
            FEBS Lett. 487 (1), 3-12 (2000)
            20584711
            PUBMED 11152876
            2 (bases 1 to 1072)
            Casaregola,S., Lepingle,A., Bon,E., Neuvéglise,C., Nguyen,H.,
            Artiguenave,F., Wincker,P. and Gaillardin,C.
            Genomic exploration of the hemiascomycetous yeasts: 7.
            Saccharomyces servazzii
            FEBS Lett. 487 (1), 47-51 (2000)
            20584717
            PUBMED 11152882
            3 (bases 1 to 1072)
            Genoscope.
            Direct Submission
            Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
            This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
COMMENT

```


Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

```
FEATURES
  source              1..1072
    /organism="Saccharomyces servazzii"
    /mol_type="genomic DNA"
    /db_xref="taxon:4311"
    /clone="XAT0AA001D07"
    /clone_lib="XAT0AA"
    /notes="end : T7"
  misc_feature        <336..>1039
    /note="similar to Saccharomyces cerevisiae ORF YLR460c [
    similarity to C.carbonum toxD protein ]
    2 putative frameshift (s)
    similar to Saccharomyces cerevisiae ORF YNL134c [
    similarity to C.carbonum toxD gene ]"
    /evidences=not_experimental
    <348..>1039
  misc_feature        /note="similar to Saccharomyces cerevisiae ORF YCR102c [
    similarity to C.carbonum toxD gene ]
    2 putative frameshift(s)"
    /evidences=not_experimental
  BASE COUNT        350 a 178 c 210 g 331 t 3 others
  ORIGIN
    Query Match      13.9%; Score 154.6; DB 29; Length 1072;
    Best Local Similarity 56.0%; Pred. No. 1.5e-27;
    Matches 394; Conservative 1; Mismatches 270; Indels 38; Gaps 4;
  QY      8 TTCCAACTCAAAAAGCGGTCTATCATTTGAAGGTGACAAAGCTGTTGTTAAAACACATG 67
  DB      337 TTCCAAAGACTATGAAGCGGTCTGTTTGTATGACAGCTGTCAAAAGTTCAAGATAGA 396
  QY      68 TCTCAGTTCAGAAATTAAGAGAGGGTACAGCTTGGTGAAGTTGAGCGCTGTTGCTGGTA 127
  DB      397 TTAAGATTCCCTTCGTGGATGATGGGTTCCTTTTAAATCAAGTTCTAGCAGTGGCAGCAA 456
  QY      128 ACCCACTGATTGGACATATTCCTTATAAGATTGGTCCAGAGTTTCAATTTCTAGGAT 187
  DB      457 ATCCAATGGATTGAAGCATGTTGAAATTTGGAATGGGTCCAAAAGGTCGAATCCTAGGTT 516
  QY      188 GTGACATTCGTGTGACAGTTGTCAAACTTGACCA---AATGCTAGTACTGACTTGAAGG 244
  DB      517 GTGATGTTCGGGTGTAGTAGTGCTATTGGTTCAGGATTCAGAGATGAATATTAAG 576
  QY      245 TTGAGATACCGGTTTCGGTTTGTTCACGGTGTCTCCCAACACAGATCTCTAAAATGGTG 304
  DB      577 TTGATGATAATGTCGTGCAATTTGTTTCATGGGGGTCAATCCGTAACCTGGCAATGGGG 636
  QY      305 CATTGCTGAATATGCCAGGGTTTATCCACCTTGTGTTT----- 343
  DB      637 CATTGACAAATTTTGTGCTGTAGATGCAATATACATTTAAATTACCTGAAACATGA 696
  QY      344 -----ACAAGATAACTTAACTCACTCAACTGCTGATGAATTTCTGAAGGCC 391
  DB      697 TGCTTCATAGCACACCGCTTTGAAACAGGTCAAGATTTTATCGAATGATGAAGGC 756
  QY      392 CTGTGAAGAACTTCGAATCTCTGTCATCATTTGCCAGTTTGTGTGACAACTGCTGGTGTGA 451
  DB      757 CAGTGGATTCTATTGAAACTGCTGCTACTATGCTGCTGGCATTAAATCACTGCAGGTGCTA 816
  QY      452 GTTTGTGTCATCACTTGGGCTCAAAAATGGAATGGACCCCATCTACCCCGCAACATCTC 511
  DB      817 TCTTAGTCCATGAATTTGGGGTTAAATTTAGAGTGGGACTGTTTGAACCCACAAATAAAT 876
  QY      512 ATCCATTATTGATTT--GGGGTGGTGTACAGCAGTGGGTCAACAACTAATCCAGTTGCC 570
  DB      877 CACAGTTTTAAATTTGGGGTGGTGGCAGCAGTAGGTCAAAATCTTAATTTCAAGTAAT 936
  QY      571 AAACATATCAATGCTTATCTAAGATTGTAAGTTGTAAGTCTTCTTAAAG--CATGAAAGCT 629
```

```
Db      937 AAAAAAATAAATCAATCTCTCAAGATTATTACAGTTGCTTCCCAAGACATGAAGATTT 996
QY      630 TTTAAAGTCTTATGGTGTGCTGATGATGCTTTTGTACTATCATGAT 672
Db      997 GCTACATCATTAATGGTGCAGATATAATATATGATTATCATGAT 1039

RESULT 5
AZ931755/c
LOCUS      AZ931755.1
DEFINITION 474.dhz89f04.s1 Saccharomyces unisporus NRRL Y-1556 Saccharomyces
            unisporus genomic clone 474.dhz89f04.s1, genomic survey sequence.
VERSION    AZ931755
KEYWORDS   GSS.
SOURCE     Saccharomyces unisporus
ORGANISM   Saccharomyces unisporus
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE  1 (bases 1 to 488)
AUTHORS   Clifton,P.F., Hillier,L.W., Fulton,L., Graves,T., Miner,T., Gish
            W.R., Waterston,R.H. and Johnston,M.
TITLE     Surveying Saccharomyces genomes to identify functional elements by
            comparative DNA sequence analysis
JOURNAL    Unpublished
COMMENT    Contact: Johnston M
            Department of Genetics
            Washington University Medical School
            Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA
            Tel: 314 362 2735
            Fax: 314 362 7855
            Email: mj@genetics.wustl.edu
            Class: random plasmid subclone.
            Location/Qualifiers
FEATURES   1..488
            /organism="Saccharomyces unisporus"
            /mol_type="genomic DNA"
            /strain="NRRL Y-1556 (CBS 398)"
            /db_xref="taxon:27294"
            /clone="474.dhz89f04.s1"
            /clone_lib="Saccharomyces unisporus NRRL Y-1556"
            /notes="Random genomic sequence"
BASE COUNT 154 a 97 c 79 g 158 t
ORIGIN
  Query Match      13.2%; Score 146.6; DB 28; Length 488;
  Best Local Similarity 58.2%; Pred. No. 1.1e-25;
  Matches 279; Conservative 0; Mismatches 194; Indels 6; Gaps 1;
  QY      256 GGTTCGGTTCGTTTCACGGTGTCTCCCAACAGATCCTAAAATGTCATTTGCTGAA 315
  DB      482 GTTCTGCTATTGTTAATGTTGTTGTCACCTGAATTCAGATTAATGGCGTTTTCGTA 423
  QY      316 TATGCCAGGGTTTATCCACCTTTGTTTTCACAG-----AGTAACCTTAACCTCACTCACT 369
  DB      422 TAGCTGAAGTGGATTTCGAATACTGCTTCAAGTTCCTCCACAAAATTAATCTTCTTAAC 363
  QY      370 GCTGATGAATTTCTGAAGCCCTCTGAAGAACTTCGAATCTGCTGTCATCAITTCGCAATT 429
  DB      362 AAGATAATATTCGAGGACGATAAATTTGACACATTTGGGAAGCTTTGTGTAGTTTCCCAATG 303
  QY      430 TCCTTGACAACTGCTGCTGTTAGTTTGTGTCATCACTTGGGCTCAAAAATGGAATGGCAC 489
  DB      302 GTATTGTATACCGCTTTAGCTGCTGTTTCTACGAATTAATAATTTGAATGGAATGCAA 243
  QY      490 CCATCTACCCCGCAACATCTCATCTCATTTTCATTTGCGGTGCTGTACAGCAGTGGGT 549
  DB      242 CCAGAAAACCAACAAATGATCATCAATCTTAATTTGGGGTGGTGTCTACTGTTTGGT 183
  QY      550 CAACTAATCAAGTTGCCAAACATA TCAATGCTTTATATAAGATTGTAAGTTGTAAGTTGCT 609
  DB      182 CAATATGCTATTCAA TTTGTCGTCAGCAAGTTTCATGGATTTTACCAGAAATTTATTACAGTGGCT 123
```

```

610 TCTAAAGGATGAAAGCTTTAAAGTCTTATGCTGCTGATGATGCTTTGACTATCAT 669
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 122 TCAAGAGAAACATGAAGAGATGTTAAATCTTTGGTGCAGATGAAGTTCGATTATCAT 63

670 GATCAGCGGTTATTGAGCAGATCAAAATCGAAGTATCAAACTGCAACCTGCAACATGTTATTGA 728
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62 GACAAGGATGTCGTGGATCAAAATTAACCAAGAAATATGATAATAAAGAACCTTAATTGA 4

RESULT 6
CNS06LW5/c
LOCUS
DEFINITION
T3 end of clone XAT0AA002A04 of library XAT0AA from strain CBS 4311
of Saccharomyces servazzii, genomic survey sequence.
ACCESSION
AL404715
VERSION
AL404715.1 GI:12166529
KEYWORDS
GSS.
SOURCE
Saccharomyces servazzii
ORGANISM
Saccharomyces servazzii
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE
1 (bases 1 to 1118)
Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G.,
Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,
de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B.,
Malpertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S.,
Saurin, W., Tekata, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,
Wincker, P. and Weissenbach, J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)
20584711
PUBMED
11152876
REFERENCE
2 (bases 1 to 1118)
Casaregola, S., Lepingle, A., Bon, E., Neuveglise, C., Nguyen, H.,
Artiguenave, F., Wincker, P. and Gaillardin, C.
Genomic exploration of the hemiascomycetous yeasts: 7.
Saccharomyces servazzii
FEBS Lett. 487 (1), 47-51 (2000)
20584717
PUBMED
11152882
REFERENCE
3 (bases 1 to 1118)
Genoscope.
Direct Submission
Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
Location/Qualifiers
1. .1118
/organism="Saccharomyces servazzii"
/mol_type="genomic DNA"
/strain="CBS 4311"
/db_xref="taxon:27293"
/clone="XAT0AA002A04"
/clone_lib="XAT0AA"
/note="end : T3"
complement(<136..>1107)
/note="similar to Saccharomyces cerevisiae ORF YNL134c [
similarity to C.carbonum toxD gene ]
2 putative frameshift(s)"
/evidence=not experimental
complement(<136..>1104)
misc_feature
misc_feature

```

```

/note="similar to Saccharomyces cerevisiae ORF YCR102c [
similarity to C.carbonum toxD gene ]
2 putative frameshift(s)"
/evidence=not experimental
BASE COUNT 366 a 217 c 202 g 328 t 5 others
ORIGIN
Query Match 12.3%; Score 137.2; DB 29; Length 1118;
Best Local Similarity 54.1%; Pred. No. 3.3e-23;
Matches 379; Conservative 1; Mismatches 294; Indels 26; Gaps 4;

QY 11 CAACCTCTCAAAAGCCGTCATCATTTGAAGGTGACAAAGCTGTTGTTAAACAGATGTCT 70
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1112 CAACCTCTATGAAGCCACCGTTATCGACAATGTAAATCTCTATCAAAAC---TGTGC 1056

QY 71 CAGTTCAGAAATTAAGAGGGGTACAGCCTTGGTGAAGGTGAGGCTGTTGCTGGTAACC 130
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1055 CATTGCCCAACTGGACGACGATATGCTTTAACACGAGCAAAATCAGTTGCTCGTAATC 996

QY 131 CAACTGATTGGAAGCATATTTGCTTATAAGATTGGTCCAGAAGGTTCAATTTCTAGGATGTG 190
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 995 CGACGGGTTGAACACCTAGCAAGTATGCTCTTGGGCCAATTTGGTTCGGTATTTGGTGTG 936

QY 191 ACATTGCTGTGTACAGTTGTCAAACTTTGGACCAAAATG---CTAGTACTGACTTGAAGGTG 247
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 935 ATCCCGCTGTGAAATTTGCAATTTAGCTTCTAAAGTCTATACACTAAATTCACCTAG 876

QY 248 GAGATACCGGTTTCGGTGTGTTTTCACGGTGTCTTCCAAACAGATCTCTAAANAATGGTGCAT 307
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 875 GTGACTGTGTATGTATGTTTATTCATGCGTCTTCCAGAGACGCTCTCGAACCGTGCAT 816

QY 308 TTGCTGAATATGCCAGGGTTTATCCACCTTTGTTTTCACAGAGTAACCTTAACCTCACTCA 367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 815 TTGCCAGTACGCTGCTGTATTTCAAGATCTATTATAA----- 774

QY 368 CTGCTGATGAAATTTCTGAAGGCCCTGTGAAGAACTTCGAATCTGCTGCATCATTTGCCAG 427
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 773 -TGAATAAGGGGTTTTTCAGTGAAGTTCCTCCTCGGAAGGTGCTGCTGCATCCCAT 715

QY 428 TTTGCTGTGACACTGCTGTTGTTAGTTTGTGTCATCCTTGGGCTCAAAAATGGAATGGC 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 714 CGTCTTTGGTTGACAGCTCTCTGCAACTCTMTTTTCT-ACACGTTCTTTAAAAAGGAATGGC 656

QY 488 ACCCATCTACCCCGCAACATACATCTCCATTATTGATTGTTGGGTGCTGTACAGCAGTGG 547
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 655 AACGAGAGAGGCCCAATAATAAAGGTACTGTTTTTMAITTTGGGATGGTCTACTGCTTGT 596

QY 548 GTCAACAACTAATCCAAAGTTGCCAAACATATCAATGCTTATATAAGATTGTAACCTGTTG 607
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 595 CTCTAGCCATTTTACAATTTATCCAAAATTCATGTTGCTGAGAAAATTTGTTGCTGTTG 536

QY 608 CTTCTAAAAGCATGAAAAGCTTTTAAAGTCTTATGCTGTGATGATGCTTTGACTATC 667
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 535 CTTGGAAGAAACGTGAAACTTAATCTTAAAGAAATTTGGGTGCTCACTGAAGTATTTGATTAT 476

QY 668 ATGATCAGCGCTTATTGAGCAGATCAAAATCGAAGTATCC 707
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 475 ATGATGCTGATATTGTTGAACAAATCAGTTCCAAATATATAC 436

RESULT 7
LOCUS
AZ9311754
DEFINITION
474.dhz89f03.s1 Saccharomyces unisporus NRRL Y-1556 Saccharomyces
unisporus genomic clone 474.dhz89f03.s1, genomic survey sequence.
ACCESSION
AZ9311754
VERSION
AZ9311754.1 GI:13502665
KEYWORDS
GSS.
SOURCE
Saccharomyces unisporus
ORGANISM
Saccharomyces unisporus
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE
1 (bases 1 to 434)

```

AUTHORS Clifton, P.F., Hillier, L.W., Fulton, L., Graves, T., Miner, T., Gish, W.R., Waterston, R.H. and Johnston, M.
TITLE Surveying Saccharomyces genomes to identify functional elements by comparative DNA sequence analysis
JOURNAL Unpublished
COMMENT Contact: Johnston M
 Department of Genetics
 Washington University Medical School
 Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA
 Tel: 314 362 2735
 Fax: 314 362 7855
 Email: mj@genetics.wustl.edu
 Class: random plasmid subclone.
FEATURES Location/Qualifiers
 source
 1. 434
 /organism="Saccharomyces unisporus"
 /mol_type="genomic DNA"
 /strain="NRRL Y-1556 (CBS 398)"
 /db_xref="taxon:27294"
 /clone="474.dhz89f03.s1"
 /clone_lib="Saccharomyces unisporus NRRL Y-1556"
 /note="Random genomic sequence"
 BASE COUNT 133 a 76 c 88 g 137 t
 ORIGIN

Query Match 11.9%; Score 132.6; DB 28; Length 434;
 Best Local Similarity 59.7%; Pred. No. 3.4e-22;
 Matches 244; Conservative 0; Mismatches 159; Indels 6; Gaps 1;
 QY 256 GGTTTCGGTTTGTTCACGGTCTCCCAACAGATCTTAAATGTGTGCTGAA 315
 DB |||||
 QY 26 GGTTCGTCTGTGATTATGTTGATGTTCAATTCAGATATGTTCTTAA 85
 DB |||||
 QY 316 TATGCCAGGGTTTATCCACCTTGTGTTTACAA-----GAGTACTTAATCACTCACT 369
 DB |||||
 QY 86 TATGTTAAAGTGGACCAACACTATTTTCAATTCCTCAAGACATTAATTTACTAAA 145
 DB |||||
 QY 370 GCTGATGAAATTTCTGAAGGCCCTGTGAAGAACTTCGAATCTGTCATCATTCGCCAGTT 429
 DB |||||
 QY 146 AAGACTGACGTGGCAACAGCAAGTTCGACATGCGAAGCTTGTGCCAGTTCCCAATG 205
 QY 430 TCGTTGCAACTGCTGTGTAGTTGTGTCATCACTTGGGCTCAAAATAGGAATGGCAC 489
 DB |||||
 QY 206 GTGTCTTACACTGTTAGCTGTACCTTTTACCAATTTGAAATTTAAATTTAGATGCAA 265
 QY 490 CCATCTACCCGCAACATACTCACTTATGATTTGGGTGGTGTACAGCAGTGGT 549
 DB |||||
 QY 266 CCAGAAGAGCCACAAATATGACCAATCTTAATTTGGGTGGTGTGCACTGGTGTGGT 325
 QY 550 CAACAACCTAATCCAAAGTTGCCAAACATATCAATGCTTATCTAAGATTTGAACCTGTTGCT 609
 DB |||||
 QY 326 CAATATGCTATCAATTTGTTAAACAAGTCCATGGTTTACTAGATATTTGTTGTTGCC 385
 QY 610 TCTAAAAGCATGAAAGCTTTTAAAGTCTTATGGTGTGATGATGCT 658
 DB |||||
 QY 386 TCTAAGAAACATGAAAGTTGTTAAATCTTATGCTGTGATGAAATCT 434

RESULT 8
AQ501781/c
LOCUS V14B11 mTn-3xHA/lacZ Insertion Library Saccharomyces cerevisiae
DEFINITION genomic 5', genomic survey sequence.
ACCESSION AQ501781
VERSION AQ501781.1 GI:4707431
KEYWORDS GSS.
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 454)
AUTHORS Ross-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A.,
 deStages, S.A., Cheung, K.-H., Sheehan, A., Symoniat, D., Jansen, R.,

Unansky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R.,
 Hager, K., Miller, P., Roeder, G.S. and Snyder, M.
TITLE Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
 Gene Disruption
JOURNAL Unpublished
COMMENT Contact: Kumar A
 Dept. of Mol. Cell. and Dev. Biology
 Michael Snyder
 Yale University
 P.O. Box 208103, New Haven, CT 06520-8103, USA
 Tel: 203 432 9949
 Fax: 203 432 6161
 Email: anuj.kumar@yale.edu
 te of mTn-3xHA/lacZ insertion.
 Seq primer: GGCCTTCTTTTGGAGTAC
 Class: transposon-tagged.
FEATURES Location/Qualifiers
 source
 1. 454
 /organism="Saccharomyces cerevisiae"
 /mol_type="genomic DNA"
 /db_xref="taxon:4932"
 /lab_host="E. coli"
 /clone_lib="mTn-3xHA/lacZ Insertion Library"
 /note="Vector: pHS56-Sal; A yeast genomic DNA library
 (lacking mitochondrial DNA) was prepared in pHS56-Sal, a
 genomic DNA was size-fractionated (DNA of roughly 2-3 kb
 in length) prior to cloning. This library was
 subsequently mutagenized with a mTn-3xHA/lacZ
 minitransposon containing lacZ, URA3, and tet resistance."
 BASE COUNT 122 a 106 c 102 g 124 t
 ORIGIN

Query Match 11.4%; Score 126.6; DB 28; Length 454;
 Best Local Similarity 64.7%; Pred. No. 1.1e-20;
 Matches 205; Conservative 0; Mismatches 109; Indels 3; Gaps 1;
 QY 4 TCAGTTCCAACCACTCAAAAGCCGTCATCATTTGAAGGTGACAAAGCTTGTGTAAACA 63
 DB |||||
 QY 401 TCGATTCCAGAAACCATGAAAGCCGTTGTCATTGAAATGGCAAGCGTGTAGTCAACAG 342
 QY 64 GATGCTCTCAGTTCAGATTTAAAGGAGGTACAGCCTTGGTGAAGTTGAGGCTGTGTGCT 123
 DB |||||
 QY 341 GACATTCCTAATTCCTGAATTTAGAAAGAGATTTGTTCTAATTAAGACTGTGCGGTTGCC 282
 QY 124 GGTAAACCAACTGATTTGGAAGCATATTTGCTTATAAGATTGGTCCAGAAGGTTCAATTTCTA 183
 DB |||||
 QY 281 GGTAAACCTTACCATTTGGAACATATTTGATTTCAAGATTGGTCTCTCAAGGTGCTCTTA 222
 QY 184 GGATGTGACATTCCTGCTACAGTTGTCAAACTTTGGACCAAAATGCTAGTACTG---ACTTG 240
 DB |||||
 QY 221 GGCTGTGATGACGCGGCCAAATCGTAAAGTTGGGCCAAATGTTGATGCTGCACGCTTT 162
 QY 241 AAGGTTGGAGATCCGGTTTCGGTTTGTTCACGGTCTTCCCAACAGATCTCTAAAT 300
 DB |||||
 QY 161 GCCATTGGGATTTACATTTATGGGTTATTTCACGGTCTTCAGTGAGGTTCCCTCAAC 102
 QY 301 GGTGCAATTTGCTGAATA 317
 DB |||||
 QY 101 GGTGCTTTGCTGAGTA 85

RESULT 9
AQ873738/c
LOCUS AQ873738
DEFINITION V7483 mTn-3xHA/lacZ Insertion Library, strain AB972 Saccharomycetes
 cerevisiae genomic 5', genomic survey sequence.
ACCESSION AQ873738
VERSION AQ873738.1 GI:6285982
KEYWORDS GSS.
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 546)

AUTHORS Ross-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A., Desbates, S.A., Cheung, K.-H., Sheehan, A., Symoniatidis, D., Jansen, R., Unanysky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R., Hager, K., Miller, P., Roeder, G.S. and Snyder, M.

TITLE Large-Scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption

JOURNAL COMMENT Unpublished

Contact: Kumar A
Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
Yale University
P.O. Box 208103, New Haven, CT 06520-8103, USA
Tel: 203 432 9949
Fax: 203 432 6161
Email: anuj.kumar@yale.edu
te of mtn-3xHA/lacZ insertion.
Seq primer: GGCCTCTCTTTGGAGTAC
Class: transposon-tagged.

```

features
  source
    1. .546
      /organism="Saccharomyces cerevisiae"
      /mol_type="genomic DNA"
      /strain="AB972 - trp1 r(0) (S288C background)"
      /db_xref="taxon:4932"

```

RESULT 10	
AQ873797/c	
LOCUS	537 bp DNA linear GSS 08-NOV-1999
DEFINITION	V75E1 Mtn-3xHA/lacZ Insertion Library, strain AB972 <i>Saccharomyces cerevisiae</i> genomic 5', genomic survey sequence.
ACCESSION	AQ873797
VERSION	AQ873797
KEYWORDS	AQ873797.1 GI:6286041
SOURCE	GSS.
ORGANISM	<i>Saccharomyces cerevisiae</i> (baker's yeast)
	<i>Saccharomyces cerevisiae</i>

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.	
REFERENCE	
1 (bases 1 to 537)	
AUTHORS	Ross-Macdonald, P., Roemer, T., Coelho, P. S. R., Agarwal, S., Kumar, A., deStaggs, S. A., Cheung, K.-H., Sheehan, A., Symoniatis, D., Jansen, R., Umansky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R., Hager, K., Miller, P., Roeder, G. S. and Snyder, M.
TITLE	Large-Scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption
JOURNAL	Unpublished
COMMENT	Contact: Kumar A Michael Snyder, Dept. of Mol. Cell. and Dev. Biology Yale University P.O. Box 208103, New Haven, CT 06520-8103, USA Tel: 203 432 9949 Fax: 203 432 6161 Email: anuj.kumar@yale.edu te of mtm-3xHA/lacZ insertion. Seq primer: GGCGTTCTTTCTTTGGAGTAC Class: transposon-tagged.

```

FEATURES
    source
        1. .537
            Location/Qualifiers
                /organism="Saccharomyces cerevisiae"
                /mol_type="Genomic DNA"
                /strain="AB972 - trp1 r(0) (S288C background)"
                /db_xref="taxon:4932"
                /lab_host="E. coli"
                /clone_lib="mtN-3xHA/lacZ Insertion Library, strain AB972"
                /notes="vector: pHS36-Sal; A yeast genomic DNA library was
                prepared in pHS36-Sal; genomic DNA was size-fractionated
                (DNA of roughly 2-3 kb in length) prior to cloning. This
                library was subsequently mutagenized with a mtN-3xHA/lacZ
                minitransposon containing lacZ, URA3, and tet resistance."

```

[illegible]

KEYWORDS Saccharomyces cerevisiae (baker's yeast)

SOURCE Saccharomyces cerevisiae

ORGANISM Saccharomycetales; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

REFERENCE 1 (bases 1 to 529)

AUTHORS Ross-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A., deBates, S.A., Cheung, K.-H., Sheehan, A., Symoniat, D., Jansen, R., Umansky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R., Heger, K., Miller, P., Roeder, G.S. and Snyder, M.

TITLE Large-Scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption

JOURNAL Unpublished

COMMENT Contact: Kumar A
Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
Yale University
P.O. Box 208103, New Haven, CT 06520-8103, USA
Tel: 203 432 9949
Fax: 203 432 6161
Email: anuj.kumar@yale.edu
te of mTn-3xHA/lacZ insertion.
Seq primer: GGCCTCTCTTCTTGGAGTAC
Class: transposon-tagged.

FEATURES
source
Location/Qualifiers
1..529
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/db_xref="taxon:4932"
/lab_host="E. coli"
/clone_lib="mTn-3xHA/lacZ Insertion Library"
/notes="Vector: pHS56-Sal; A yeast genomic DNA library (lacking mitochondrial Sal) was prepared in pHS56-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mTn-3xHA/lacZ minitransposon containing lacZ, URA3, and tet resistance."
BASE COUNT 152 a 117 c 112 g 147 t 1 others

Query Match 11.1%; Score 124; DB 28; Length 529;
Best Local Similarity 64.0%; Pred. No. 5.1e-20;
Matches 203; Conservative 0; Mismatches 111; Indels 3; Gaps 1;

Qy 4 TCAGTTCACCACTCAAAAGCCGTCATCATTAAGAGTGACAAAGCTGTGTGTTAAACA 63
Db 415 TCGATTCCAGAAACCATGAAAGCGTGTCTCATTTGAAATGCGAGGCTGTAGTCANACAG 356

Qy 64 GATGCTCAGTTCAGAAATTAAGAGGGTACAGCTTGGTGAAGGTTGAGGCTGTGCT 123
Db 355 GACATTCCAATTCCTGAATTAAGAAGAGGATTTGTTCTTAATTAAGACTGTGCGCGTAGCC 296

Qy 124 GGTAACCAACTGATTCGAACCATATTGCTTATAGATTGGTCCAGAGGTTCAATTCTA 183
Db 295 GGTAACCTCCGATTCGAACCATATTGATTTCAAGATTGGTCTCAAGGTGCGCTCTTA 236

Qy 184 GGATGTGACATTTGCTGTACAGTTGTCAAACTGGACCAATGCTAGTACTG---ACTTG 240
Db 235 GGCTGTGATGACGCGGCCCAATGTAAGTTGGCCCAATGTTGATGCTGACGCTTT 176

Qy 241 AAGTTGGAGATACCGTTTCGGTTTGTTCACGGTGCTTCCCAACAGATCTCAAAAAT 300
Db 175 GCCATTGGTATTACATTTATGGGGTTATTACCGTGTCTTACCGTGTCTTCCCTCAAAC 116

Qy 301 GGTCGATTTGCTGAATA 317
Db 115 GGTGCTTTGCTGAGTA 99

RESULT 12
BZ300244/c
LOCUS BZ300244 467 bp DNA linear GSS 31-OCT-2002
DEFINITION KD0613.R1 Kluyveromyces delphensis Random Genomic Library
Kluyveromyces delphensis genomic clone KD0613, genomic survey

sequence.
BZ300244
VERSION BZ300244.1 GI:24444834
KEYWORDS GSS.
SOURCE Kluyveromyces delphensis
ORGANISM Kluyveromycetes delphensis
REFERENCE 1 (bases 1 to 467)
AUTHORS Wong, S., Fares, M.A., Zimmermann, W., Butler, G. and Wolfe, K.H.
TITLE Evidence from comparative genomics for a complete sexual cycle in the 'asexual' pathogenic yeast *Candida glabrata*

JOURNAL Genome Biol. 4 (2), R10 (2003)

MEDLINE 22508158
PUBMED 12620120

COMMENT Contact: Wong S
Department of Genetics, Smurfit Institute
Trinity College Dublin
Dublin 2, Ireland
Tel: 353 1 6082319
Fax: 353 1 6798558
Email: swong@tcd.ie
Class: plasmid ends.

FEATURES
Location/Qualifiers
1..467
/organism="Kluyveromyces delphensis"
/mol_type="genomic DNA"
/strain="CBS 2170"
/db_xref="taxon:51657"
/clone_lib="KD0613"
/clone_lib="Kluyveromyces delphensis Random Genomic Library"

BASE COUNT 118 a 108 c 118 g 123 t

Query Match 11.1%; Score 123.8; DB 29; Length 467;
Best Local Similarity 57.8%; Pred. No. 5.5e-20;
Matches 262; Conservative 0; Mismatches 182; Indels 9; Gaps 2;

Qy 60 AACAGATGCTCAGTTCAGAAATTAAGAGGGGTACAGCCCTTGGTGAAGCTGAGGCTGT 119
Db 461 AAAGGAGGTTCTTTCGACCAATGGGTGAAACCAATGTTCTGATTAAAGAACACTGCGGT 402

Qy 120 TGCTGTAACCCCACTGATTGGAAGCATATTGCTTAAAGATTGGTCCAGAAGGTTCAAT 179
Db 401 TGCTGTAACCCCACTGACTGGAAGCATCGCATACAGATCGGTCCACAGGATCAT 342

Qy 180 TCTAGGATGTGACATTTGCTGTACAGTTGTCAAACTTGGACCAAAAT---GCTAGTACTGA 236
Db 341 CTGTGGCTGTGACGCTACTGCTGAGATTGTCAAGTTAGGTGCTGGTGTCTTCCAA 282

Qy 237 CTTGAAGGTTGGAGATACCGGTTTCGGTTTTGTTTCACGGTGTCTCCAAACAGATCTTAA 296
Db 281 GTACAAGCTCGGTCAAAAGGTCTACGGTTTTGGTCCACGGTTCCTCCGTCGAAGCACCTCA 222

Qy 297 AAATGGTGCAATTTGCTGAATATCCAGGGTTTATCCACCTTTGTTTTAC-----AAGAG 350
Db 221 AAACGGTGCCTTCGAGAAATTTCTGTTTTGGATCTCTAAAGTCATGTACGTTCCAAACCA 162

Qy 351 TAACTTAACCTCACTCAACTGCTGATGAAATTTCTGAAGGCCCTGTGGAAGAACTTCGAATC 410
Db 161 AGATATTACCATCAGCGGTAAGGACGAGATCCAGCTGGTCCAGTTAAGCATGTTGAAGA 102

Qy 411 TGCTGCATCATTCGACGTTTCGGTTTGACAACTGCTGGTGTAGTTGTGTATCATCTATGGG 470
Db 101 CACAGCAACATTCACAGTCTCTGTTGACCACTGCTGGTACTGTTTTGACACCAATTTCCG 42

Qy 471 CTCAAAAATGGAATGGCACCCATCTACCCGCA 503
Db 41 TCTAGATTGCTCTGGGAACCATCCAGGCCACA 9

RESULT 13

CNS06HG8 946 bp DNA linear GSS 30-NOV-2001
 LOCUS T7 end of clone AS0AA012C02 of library AS0AA from strain CLIB 533
 of Saccharomyces bayanus, genomic survey sequence.

ACCESSION AL398958
 VERSION AL398958.1 GI:12152890
 KEYWORDS GSS.

ORGANISM

Saccharomyces bayanus
 Saccharomyces bayanus
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

REFERENCE

1 (bases 1 to 946)
 Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G.,
 Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,
 de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B.,
 Malpertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S.,
 Saurin, W., Tokala, P., Toffano-Nioche, C., Wesolowski-Louvel, M.,
 Wincker, P. and Weissenbach, J.
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of
 yeast species for molecular evolution studies

TITLE

Genomic exploration of the hemiascomycetous yeasts: 1. A set of
 yeast species for molecular evolution studies

JOURNAL

FEBS Lett. 487 (1), 3-12 (2000)

REFERENCE

1152876
 2 (bases 1 to 946)
 Bon, E., Neuveglise, C., Casaregola, S., Artiguenave, F., Wincker, P.,
 Aigle, M. and Durrens, P.
 Genomic exploration of the hemiascomycetous yeasts: 5.
 Saccharomyces bayanus var. uvarum

JOURNAL

FEBS Lett. 487 (1), 37-41 (2000)

REFERENCE

1152880
 3 (bases 1 to 946)
 Genoscope.

AUTHORS

Direct Submission

JOURNAL

Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
 2 rue Gaston Cremieux, CP 7506, 91057 EVRY cedex, FRANCE. (E-mail :
 seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT

This GSS is part of a random genomic sequencing program of thirteen
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
 exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,
 Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
 lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
 angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
 5 kb were prepared and both extremities were sequenced. See
 keywords for description of this sequence and for the sequence of
 the other extremity of this insert.

FEATURES

source
 1..946
 /organism="Saccharomyces bayanus"
 /mol_type="genomic DNA"
 /strain="CLIB 533"
 /variety="uvarum"
 /db_xref="taxon:4931"
 /clone="AS0AA012C02"
 /clone_lib="AS0AA"
 /note="Tend : T7"
 misc_feature
 <485..>944
 /note="similar to Saccharomyces cerevisiae ORF YNL134c {
 similarity to C.carbonum toxD gene }
 2 putative frameshift(s)"
 /evidence=not_experimental
 misc_feature
 <523..>944
 /note="similar to Saccharomyces cerevisiae ORF YCK102c {
 similarity to C.carbonum toxD gene }
 1 putative frameshift(s)
 similar to Saccharomyces cerevisiae ORF YLR460c {
 similarity to C.carbonum toxD protein }"
 /evidence=not_experimental

BASE COUNT

252 a 219 c 200 g 271 t 4 others

ORIGIN

10.7%; Score 119; DB 29; Length 946;

Query Match

10.7%; Score 119; DB 29; Length 946;

Best Local Similarity 59.8%; Pred. No. 1.le-18;
 Matches 253; Conservative 1; Mismatches 161; Indels 8; Gaps 3;

Qy 32 TCATTGAAGTGAACAAGCTGTTGTTAAACACATGCTCAGTTCAGAGTAATTAAGGAGG 91
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 524 TTATTGAAGCGGATGAAGCTGTCTATCAAGGAGGGTGTGCCCATTCGCCAATTAGAGGATG 583
 Qy 92 GTACAGCCTTGGTGAAGGTTGAGGCTGTTGCTGTAACCCAACTGATTGGAAGCATATTG 151
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 584 GTTTCGTTCTGATCAAAACCGTGCCTGTTGCCGGTAAATCCACTGATTGGGCACATATTG 643
 Qy 152 CTTATAAGATTTGTTCCAGAGGTTCAATTTCTAGGATGTGACATTTGCTGGTACAGTTGTCA 211
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 644 ACTACAAGATAGAGCCCTCAAGGTTCTATCTCTGGTGGCGATGCTACAGGTCCAATTGTTA 703
 Qy 212 AACTTGGACCAAT---GCTAGTACTGACTTGAAGGTTGGAGATACGGTTTCGGTTTGG 268
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 704 AACTGGGTCGGGTGTGATGGCAACGCTTTGCCATTGGAGATTACATTTACGGGTTTA 763
 Qy 269 TTCACGGTGTCTCCCAACACAGATCCTAAAAATGGTCATTTGCTGAATATGCCAGGTTT 328
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 764 TTCACGGTGTCTCCGAGATTCCCTCCAAUGGTCCTTCGCTGATATCTGCAATTC 823
 Qy 329 ATCCACCTTTGTTTATCAA---GAGTAACCTTAACCTCACTCAACTGCTGATGAAA--TTTC 383
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 824 CATCGGAGATGCCATACAAATCATCAACGATATTAAGTTGTGTGTAAGGATAGTCTAC 883
 Qy 384 TGAAGGCCCTGTGAAGAACTTCGAATCTGCTGCATCATTTGCCAGTTCTGTTGCAACTGC 443
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 884 CTGAAGTCCCGTAAAGTCTTTTAGAAGGCGCTGTCACTATTTCGGTGCACATTGACMACTGC 943
 Qy 444 TGG 446
 Db 944 CGG 946

RESULT 14
 A0873736/c
 LOCUS 542 bp DNA linear GSS 08-NOV-1999
 DEFINITION V74E1 mtn-3xHA/lacZ Insertion Library, strain AB972 Saccharomyces
 cerevisiae genomic 5', genomic survey sequence.

ACCESSION A0873736
 VERSION A0873736.1 GI:6285980
 KEYWORDS GSS.

SOURCE Saccharomyces cerevisiae (baker's yeast)
 ORGANISM Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

REFERENCE 1 (bases 1 to 542)
 AUTHORS Ross-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A.,
 desEtages, S.A., Cheung, K.-H., Sheehan, A., Symoniatidis, D., Jansen, R.,
 Umansky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R.,
 Hager, K., Miller, P., Roeder, G.S. and Snyder, M.

TITLE Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
 Gene Disruption

JOURNAL Unpublished
 COMMENT Contact: Kumar A
 Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
 Yale University
 P.O. Box 208103, New Haven, CT 06520-8103, USA
 Tel: 203 432 9949
 Fax: 203 432 6161
 Email: anuj.kumar@yale.edu

te of mtn-3xHA/lacZ insertion.
 Seq primer: GGCCTCTTTCTTTGGAAGTAC
 Class: transposon-tagged.

FEATURES
 source
 1..542
 /organism="Saccharomyces cerevisiae"
 /mol_type="genomic DNA"
 /strain="AB972 - trp1 r(0) (S288C background)"
 /db_xref="taxon:4932"
 /lab_host="E. coli"

/clone lib="mTn-3xHA/lacZ Insertion Library, strain AB972"
/notes="Vector: pHS56-Sal; A yeast genomic DNA library was
prepared in pHS56-Sal; genomic DNA was size-fractionated
(DNA of roughly 2-3 kb in length) prior to cloning. This
library was subsequently mutagenized with a mTn-3xHA/lacZ
minitransposon containing lacZ, URA3, and tet resistance."

```
BASE COUNT      152 a   136 c   112 g   142 t
ORIGIN
Query Match      10.6%; Score 117.8; DB 28; Length 542;
Best Local Similarity 63.3%; Pred. No. 1.8e-18; Indels 3; Gaps 1;
Matches 198; Conservative 0; Mismatches 112;
QY 8 TTCCAACCACTCAAAAAGCGCTCATCTTCAAGGTGACAAAGCTGTGTTAAAACAGATG 67
   |||||
DB 409 TTCCAGAACCATGAAAGCGGTGTCATTGAAATGGCGAGGTGTAGTCAGACAGACA 350
   |||||
QY 68 TCTCAGTTCAGAAATTAAAGAGGGGTACAGCCTTGGTGAAGGTGAGGCTGTGCTGGTA 127
   |||||
DB 349 TTCCAATTCTCGCATTTAGAAAGAGGATTTGTTCTAATTAGACTGTGCGCGTTGCCGGTA 290
   |||||
QY 128 ACCCACTGATTCGAGCATATTGCTTATAAGATTGGTCCAGAGGTTCATTTCTAGGAT 187
   |||||
DB 289 ACCATACCGATTGAAACATATTGATTTTCGAGATTGGTCTCAAGGTGCCCTCTTAGGCT 230
   |||||
QY 188 GTGACATTGCTGTACAGTTGTCAAACCTTGACCAAAATGCTAGTACTG---ACTTGAAGG 244
   |||||
DB 229 GTGATGCGCGGCCAAATCGTAAAGTTGGGCCAAATGTTGATGCTGCACGCTTGCCA 170
   |||||
QY 245 TTGGAGATACCGGTTTCGGTTTTGTTTCACGGTGCTTCCCAACAGATCTTAAAAATGGTG 304
   |||||
DB 169 TTGGTGATTACATTTATGGGGTTATTTCACGGTGCTTTCAGTGAGGTTCCTCCCTCAACCGTG 110
   |||||
QY 305 CATTGCTGAATA 317
   |||||
DB 109 CATTGCTGAGTA 97
```

```
RESULT 15
LOCUS
DEFINITION EST104002 S. cerevisiae strain X2180-1A Saccharomyces cerevisiae
cDNA 5' end, mRNA sequence.
ACCESSION T38505
VERSION T38505.1 GI:622322
KEYWORDS EST.
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 337)
AUTHORS Weinstein, K.
TITLE Saccharomyces cerevisiae cDNAs
JOURNAL Unpublished
COMMENT The Institute for Genomic Research
932 Clopper Rd, Gaithersburg, MD 20878
Tel: 3018699056
Fax: 3018699423
Email: tdbinfo@db.tigr.org
For clone availability please contact the TIGR Database
(tdbinfo@db.tigr.org)
Seq primer: M13 Reverse.
```

```
FEATURES
source
1..337
/organism="Saccharomyces cerevisiae"
/mol_type="mRNA"
/strain="X2180-1A"
/db_xref="taxon:4932"
/clone_lib="S. cerevisiae strain X2180-1A"
/notes="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI"
```

```
BASE COUNT      152 a   136 c   112 g   142 t
ORIGIN
Query Match      10.6%; Score 117.8; DB 28; Length 542;
Best Local Similarity 63.3%; Pred. No. 1.8e-18; Indels 3; Gaps 1;
Matches 198; Conservative 0; Mismatches 112;
QY 8 TTCCAACCACTCAAAAAGCGCTCATCTTCAAGGTGACAAAGCTGTGTTAAAACAGATG 67
   |||||
DB 409 TTCCAGAACCATGAAAGCGGTGTCATTGAAATGGCGAGGTGTAGTCAGACAGACA 350
   |||||
QY 68 TCTCAGTTCAGAAATTAAAGAGGGGTACAGCCTTGGTGAAGGTGAGGCTGTGCTGGTA 127
   |||||
DB 349 TTCCAATTCTCGCATTTAGAAAGAGGATTTGTTCTAATTAGACTGTGCGCGTTGCCGGTA 290
   |||||
QY 128 ACCCACTGATTCGAGCATATTGCTTATAAGATTGGTCCAGAGGTTCATTTCTAGGAT 187
   |||||
DB 289 ACCATACCGATTGAAACATATTGATTTTCGAGATTGGTCTCAAGGTGCCCTCTTAGGCT 230
   |||||
QY 188 GTGACATTGCTGTACAGTTGTCAAACCTTGACCAAAATGCTAGTACTG---ACTTGAAGG 244
   |||||
DB 229 GTGATGCGCGGCCAAATCGTAAAGTTGGGCCAAATGTTGATGCTGCACGCTTGCCA 170
   |||||
QY 245 TTGGAGATACCGGTTTCGGTTTTGTTTCACGGTGCTTCCCAACAGATCTTAAAAATGGTG 304
   |||||
DB 169 TTGGTGATTACATTTATGGGGTTATTTCACGGTGCTTTCAGTGAGGTTCCTCCCTCAACCGTG 110
   |||||
QY 305 CATTGCTGAATA 317
   |||||
DB 109 CATTGCTGAGTA 97
```

```
RESULT 15
LOCUS
DEFINITION EST104002 S. cerevisiae strain X2180-1A Saccharomyces cerevisiae
cDNA 5' end, mRNA sequence.
ACCESSION T38505
VERSION T38505.1 GI:622322
KEYWORDS EST.
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 337)
AUTHORS Weinstein, K.
TITLE Saccharomyces cerevisiae cDNAs
JOURNAL Unpublished
COMMENT The Institute for Genomic Research
932 Clopper Rd, Gaithersburg, MD 20878
Tel: 3018699056
Fax: 3018699423
Email: tdbinfo@db.tigr.org
For clone availability please contact the TIGR Database
(tdbinfo@db.tigr.org)
Seq primer: M13 Reverse.
```

```
FEATURES
source
1..337
/organism="Saccharomyces cerevisiae"
/mol_type="mRNA"
/strain="X2180-1A"
/db_xref="taxon:4932"
/clone_lib="S. cerevisiae strain X2180-1A"
/notes="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI"
```

```
BASE COUNT      83 a   75 c   77 g   97 t   5 others
ORIGIN
```

Query Match 10.4%; Score 115.4; DB 14; Length 337;
Best Local Similarity 63.8%; Pred. No. 6.1e-18;
Matches 203; Conservative 0; Mismatches 111; Indels 4; Gaps 2;

```
QY 4 TCAGTTCCCAACCACTCAAAAAGCGCTCATCTTGAAGGTGACAAAGCTGTGTTAAAACA 63
   |||||
DB 2 TCAGTTCCAGAACCATGAAAGCGGTTCATTGAAATGGCAAGCGCTGTAGTCAACAG 61
   |||||
QY 64 GATGTTCTCAGTTCAGAAATTAAAGAGGGGTACAGCCTTGGTGAAGGTGTGAGGCTGTGCT 123
   |||||
DB 62 GNCATTTCGAATTCCTGAATTAGAAGAGGATTTGTTCTAATTAAAGACTGTGCGCGTTGCC 121
   |||||
QY 124 GGTAAACCCCAACTGATTGGGAAGCATATTGCTTATAAGATTGGTCCAGAAAGGTTCATTTCTA 183
   |||||
DB 122 GGTAAACCCCTACCGATTGGGAACATATTGATTTCAAGATTGGTCTCAAGGTGCGCTCTTA 181
   |||||
QY 184 GGATGTGACATTCGCTGTGATCAGTTGTCAAACCTTGGACCAAAATGCTAGTACTG---ACTTG 240
   |||||
DB 182 GGCTGTGATGTCAGCGCGGCCAAATCGTAAAGTTGGGCCAAATGTTTATGCTGCACGCTTT 241
   |||||
QY 241 AAGGTTGGAGATACCGGTTTCGGTTTTGTTTCACGGTGCTTCCCAACAGAT-CCTAAAAA 299
   |||||
DB 242 GCCATTGGTGATTACATTTATGGGGTNATTACGGTGCTTTTCAGTGAGGTTCCTCCCTCAAA 301
   |||||
QY 300 TGGTGCAATTGCTGAATA 317
   |||||
DB 302 CGGTGCTGCTGAGTA 319
```

```
RESULT 16
LOCUS
DEFINITION 476.dis09b12.s1 Saccharomyces castellii NRRL Y-12630 Saccharomyces
castellii genomic clone 476.dis09b12.s1, genomic survey sequence.
ACCESSION AZ926593
VERSION AZ926593.1 GI:13497495
KEYWORDS GSS.
SOURCE Saccharomyces castellii
ORGANISM Saccharomyces castellii
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 540)
AUTHORS Clifton, P.F., Hillier, L.W., Fulton, L., Graves, T., Miner, T., Gish,
W.R., Waterston, R.H. and Johnston, M.
TITLE Surveying Saccharomyces genomes to identify functional elements by
comparative DNA sequence analysis
JOURNAL Unpublished
COMMENT Contact: Johnston M
Department of Genetics
Washington University Medical School
Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA
Tel: 314 362 2735
Fax: 314 362 7855
Email: mj@genetics.wustl.edu
Class: random plasmid subclone.
```

```
FEATURES
source
1..540
/organism="Saccharomyces castellii"
/mol_type="genomic DNA"
/strain="NRRL Y-12630 (CBS 4309)"
/db_xref="taxon:27288"
/clone_lib="476.dis09b12.s1"
/notes="Random genomic sequence"
BASE COUNT      182 a   80 c   89 g   189 t
```

```
BASE COUNT      182 a   80 c   89 g   189 t
ORIGIN
Query Match      9.7%; Score 108.4; DB 28; Length 540;
Best Local Similarity 61.9%; Pred. No. 3.9e-16;
Matches 172; Conservative 0; Mismatches 106; Indels 0; Gaps 0;
```


genomic clone CG0601, genomic survey sequence.

ACCESSION BZ293919
VERSION BZ293919.1 GI:24434218
KEYWORDS GSS.
SOURCE Candida glabrata

ORGANISM

REFERENCE
AUTHORS Wong, S., Fares, M.A., Zimmermann, W., Butler, G. and Wolfe, K.H.

TITLE Evidence from comparative genomics for a complete sexual cycle in the 'asexual' pathogenic yeast *Candida glabrata*

JOURNAL Genome Biol. 4 (2), R10 (2003)

MEDLINE 22508158

PUBMED 12620120

COMMENT Contact: Wong S
Department of Genetics, Smurfit Institute
Trinity College Dublin
Dublin 2, Ireland
Tel: 353 1 6082319
Fax: 353 1 6798558
Email: swong@tcd.ie
Class: plasmid ends.

FEATURES

source

1. 574

/organism="Candida glabrata"

/mol_type="genomic DNA"

/strain="CBS 138"

/db_xref="taxon:5478"

/clones="CG0601"

/clone.lib="Candida glabrata Random Genomic Library"

BASE COUNT 193 a 104 c 106 g 181 t

ORIGIN

Query Match 6.6%; Score 73.2; DB 29; Length 574;

Best Local Similarity 69.7%; Pred. No. 2.4e-07;

Matches 99; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 60 AACAGATGCTCAGTCCAGATTAAGAGGGGTACAGCCTTGGTGAAGTTGAGGCTGT 119

DB 429 AAAGGAAGTTCCTGTCAGAAATTCGCGTGGTGTCTAGTTAAGAAATAGAGCTGT 488

QY 120 TGCTGGTAACCAACTGATGGAAGCATATTCCTATAGATGTCACAGAGTTTCAAT 179

DB 489 CGCTGGTAACCAACTGGAAGCATACATGGAAGTTGACACACAGTTCCAT 548

QY 180 TCTAGGATGACATTCCTGT 201

DB 549 CGCGGTTGTACATTCGCGT 570

RESULT 22

AQ501614/c

LOCUS

DEFINITION

V17F2 mtN-3xHA/lacZ Insertion Library Saccharomyces cerevisiae

genomic 5', genomic survey sequence.

ACCESSION AQ501614

VERSION AQ501614.1 GI:4707264

KEYWORDS GSS.

SOURCE

ORGANISM

Saccharomyces cerevisiae (baker's yeast)

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

1 (bases 1 to 454)

REFERENCE

AUTHORS Ross-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A.,

desEtages, S.A., Cheung, K.-H., Sheehan, A., Symoniat, D., Jansen, R.,

Unanue, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R.,

Hager, K., Miller, P., Roeder, G.S., and Snyder, M.

Large-Scale Analysis of the Yeast Genome by Transposon Tagging and

Gene Disruption

Unpublished

JOURNAL

COMMENT Contact: Kumar A

Michael Snyder, Dept. of Mol. Cell. and Dev. Biology

Yale University
P.O. Box 208103, New Haven, CT 06520-8103, USA

Tel: 203 432 9949

Fax: 203 432 6161

Email: anuj.kumar@yale.edu

te of mtN-3xHA/lacZ insertion.

Seq primer: GGCTCTCTCTTTTGAAGTAC

Class: transposon-tagged

Location/Qualifiers

1. 454

/organism="Saccharomyces cerevisiae"

/mol_type="genomic DNA"

/db_xref="taxon:4932"

/lab_host="E. coli"

/clone.lib="mtN-3xHA/lacZ Insertion Library"

/notes="Vector: pHS86-Sal; A yeast genomic DNA library

(lacking mitochondrial DNA) was prepared in pHS86-Sal;

genomic DNA was size-fractionated (DNA of roughly 2-3 kb

in length) prior to cloning. This library was

subsequently mutagenized with a mtN-3xHA/lacZ

minitransposon containing lacZ, URA3, and tet resistance."

BASE COUNT 123 a 114 c 71 g 145 t

ORIGIN

Query Match 5.5%; Score 61.2; DB 28; Length 454;

Best Local Similarity 65.2%; Pred. No. 0.00022;

Matches 90; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 8 TTCCAACCACTCAAAAAGCGTCATCTGAAGGTGACAAAGCTGTGTAAACAGATG 67

DB 177 TTCCAGAACCCAGGAAACCGTTTTCATGGAATGCAAGGCTGTAGTCAACAGGACA 118

QY 68 TCTCAGTTCCAGAAATTAAGAGGGGTACAGCCTTGTGAAGTTGAGGCTGTTCGTGTA 127

DB 117 TTCCAATTCCTGAATTAAGAAGGATTTGTTCTAATTAAGACTGTGCGCGTTGCCGTA 58

QY 128 ACCCAACTGATTGGAAGC 145

DB 57 ACCCTACCATTTGGAAC 40

RESULT 23

CNS07BB3

LOCUS

DEFINITION

T3 end of clone BC0AA010F11 of library BC0AA from strain CBS 767 of

Debaryomyces hansenii, genomic survey sequence.

ACCESSION AL437653

VERSION AL437653.1 GI:12221066

KEYWORDS GSS.

SOURCE

ORGANISM

Debaryomyces hansenii (anamorph: Candida famata)

Debaryomyces hansenii

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; Saccharomycetaceae; Debaryomyces.

1 (bases 1 to 1146)

REFERENCE

AUTHORS Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G.,

Bolotin-Fukuhara, M., Bon, E., Brotsch, P., Casaregola, S.,

de-Montigny, J., Dujon, B., Durand, F., Lepingle, A., Lorente, B.,

Malpertuy, A., Neuvéglise, C., Ozier-Kalogeropoulos, O., Potier, S.,

Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,

Wincker, P. and Weissenbach, J.

Genomic exploration of the hemiascomycetous yeasts: 1. A set of

yeast species for molecular evolution studies

FEBS Lett. 487 (1), 3-12 (2000)

JOURNAL

MEDLINE 20584711

PUBMED 11152876

REFERENCE

AUTHORS Lepingle, A., Casaregola, S., Neuvéglise, C., Bon, E., Nguyen, H.,

Artiguenave, F., Wincker, P. and Gaillardin, C.

Genomic exploration of the hemiascomycetous yeasts: 14.

Debaryomyces hansenii var. hansenii

FEBS Lett. 487 (1), 82-86 (2000)

JOURNAL

MEDLINE 20584724

PUBMED 11152889

AUTHORS Whitfield, C.W., Band, M.R., Bonaldo, M.F., Kumar, C.G., Liu, L.,
 Pardinas, J., Robertson, H.M., Soares, B. and Robinson, G.E.
TITLE Annotated expressed sequence tags and cDNA microarrays for studies
 of brain and behavior in the honey bee
JOURNAL Genome Res. 12 (4), 555-566 (2002)
MEDLINE 21929762
PUBMED 11932240
COMMENT Contact: Gene E. Robinson
 Department of Entomology
 University of Illinois
 505 S. Goodwin Ave., Urbana, IL 61801, USA
 Tel: 217 265 0309
 Fax: 217 244 3499
 Email: generobi@life.uiuc.edu
 This research was funded by the University of Illinois Critical
 Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation
 Award in Functional Genomics to G.E. Robinson and an NSF
 Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
PCR Primers
FORWARD: TAATACGACTCACTATAGG
BACKWARD: ATTAACCTCACTAAG
Plate: BBL60022A20 row: C column: 02
Seq primer: AGCGATAACAAATTCACACAGGA
High quality sequence stop: 515.

FEATURES source

1. .515
 /organism="Apis mellifera"
 /mol_type="mRNA"
 /strain="mixed strains of European bees, predominantly
 A.m. ligustica"
 /db_xref="taxon:7460"
 /clone="BBL60022A20C02"
 /sex="female"
 /tissue_type="brain"
 /dev_stage="adult worker honey bee"
 /lab_host="DH10B"
 /clone_lib="Bee Brain Normalized Library, BBL6"
 /note="Organ: brain; Vector: pT73-Pac; Site 1: EcoRI;
 Site 2: NotI; The BBL6 library was contributed by the
 Soares laboratory and it was constructed and normalized
 as described by Bonaldo, M.F., Lennon, G. and Soares,
 M.B. (1996), Genome Research 6(9): 791-806. RNA was
 prepared from dissected brains of adult worker bees of
 various ages and various behavioral groups."
 198 a 74 c 89 g 153 t 1 others

BASE COUNT ORIGIN

Query Match 4.6%; Score 51.6; DB 12; Length 515;
 Best Local Similarity 52.8%; Pred. No. 0.056;
 Matches 134; Conservative 0; Mismatches 119; Indels 1; Gaps 1;
 Qy 766 AAGTCACAGATAGTCTACCTGCCACATATTAGAGTGGTTCCAAATGACATTGAA 825
 Db 129 AAATTCATTTGATATTGCTCATACATCATATTATCAAGGGCTCAAGTACTCTTAAA 188
 Qy 826 AGCATCTCGAAGAAATCAGAAAGATGTTAAATTCATATCTTTGTTGTTATCGT 885
 Db 189 GCAATTTTAAAGAAATAGGACAAAGAAACATTACTGTTCAAAGACATGGCGATTAAAT 248
 Qy 886 GCATCTGGTCAAGAAATTTCTATTTGGGTGCAACAGATTTCTGCTAGTCCAGAAATCAT 945
 Db 249 GAACGTCATTATGAGGTTTGACTGGTTTAAATAAAGCTGAAACTGCTGCTAATATGGT 308
 Qy 946 GAAGCCACAGTT-AAATTCGTTAAGTTTATAAATCCACCTTAAACACGGTGATATCCA 1004
 Db 309 GAAGAACAAAGTTCAAATTTGGAGAAGATCTTTTGATACACCTCTCCACCTATGGAACCA 368
 Qy 1005 TCATATCAATATTA 1018
 Db 369 GATCATAAATATTA 382

RESULT 26

BP099113
LOCUS 595 bp mRNA linear EST 31-DEC-2002
DEFINITION BP099113 Chaetomium cDNA library Chaetomium globosum cDNA clone
 high chaeB.Contig224, mRNA sequence.
ACCESSION BP099113
VERSION BP099113.1 GI:27438553
KEYWORDS EST.
SOURCE Chaetomium globosum
ORGANISM Chaetomium globosum
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Sordariomycetidae; Sordariales; Chaetomiaceae; Chaetomium.
REFERENCE 1 (bases 1 to 595)
AUTHORS Yang, Q. and Jin, H.
TITLE Chaetomium mycelium EST
JOURNAL Unpublished
COMMENT Contact: Qian Yang
 Department of Biotechnology
 Harbin Institute of Technology
 West dazhi 92, Harbin, Heilongjiang 150001, China
 Tel: 86-451-563-5236
 Fax: 86-451-641-2952
 Email: kimhongseng@x263.net.

FEATURES source

1. .595
 /organism="Chaetomium globosum"
 /mol_type="mRNA"
 /db_xref="taxon:38033"
 /clone="high chaeB.Contig224"
 /dev_stage="mycelium"
 /clone_lib="Chaetomium cDNA library"
 125 a 214 c 163 g

BASE COUNT ORIGIN

Query Match 4.6%; Score 51.6; DB 12; Length 595;
 Best Local Similarity 47.5%; Pred. No. 0.058;
 Matches 153; Conservative 0; Mismatches 169; Indels 0; Gaps 0;
 Qy 1 ATGTCAGTTCCACCACTCAAAGCCGTCATCATTTGAAGGTGACAAAGCTGTTGTATA 60
 Db 57 ATGCGCCCTCCATCAGAGATCAATGCGGTGTCATCACCAGCCCGAAGCGTGAAGTC 116
 Qy 61 ACAGATGTCTCAGTTCCAGAAATTAAGAGAGGGGTACAGCCTTGTGAAGGTTGAGGCTGT 120
 Db 117 AAGACGGTACCCCTCCAGAGCTCCCGGATGACTACATCTCTCGCCGACCGGCGTG 176
 Qy 121 GCTGGTAACCCAACTGATTGGAAGCATATGCTTATAAGATTGGTCCAGAGGTTCAATT 180
 Db 177 GCTCTCAACCCGACAGACTGGAAGCAGCTTCAGTGGGGGTGAGCACCGTAGGACCCGT 236
 Qy 181 CTAGGATGTGACATTCCTGCTACAGTTGTCAAACTTGGACCAATGCTAGTACTGACTTG 240
 Db 237 GTAGGCTGGGACTACCGCGCATCGTCGAGAGTGGGGCCCAAGTGAACAGCCCTTC 296
 Qy 241 AAGTTTGGAGATACCGGTTTCGGTTTGTTCACGGTGTCTCCCAACAGATCTCTAAAAAT 300
 Db 297 AAAAAGGTTGACCGTATCTCGGAATTGTCCACGGAGCCAACTCGCTGCGGCTGATGCG 356
 Qy 301 GGTGCAATTTGCTGAATATGCCA 322
 Db 357 GGTGGTTTTCGCGACTATACATCA 378

RESULT 27 B1505803 **LOCUS**

DEFINITION B1505803 B170024A10H01.5 Bee Brain Normalized/Subtracted Library, B17 Apis
 mellifera cDNA clone B170024A10H01 5', mRNA sequence.
ACCESSION B1505803
VERSION B1505803.1 GI:15356177
KEYWORDS EST.
SOURCE Apis mellifera (honeybee)
ORGANISM Apis mellifera
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea; Apidae; Apis.

1 (bases 1 to 619)

Whitfield, C.W., Band, M.R., Bonaldo, M.F., Kumar, C.G., Liu, L., Pardinas, J., Robertson, H.M., Soares, B. and Robinson, G.E.

Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee

Genome Res. 12 (4), 555-566 (2002).

21929762

11932240

Contact: Gene E. Robinson
Department of Entomology
University of Illinois
505 S. Goodwin Ave., Urbana, IL 61801, USA
Tel: 217 265 0309
Fax: 217 244 3499
Email: generobi@life.uiuc.edu

This research was funded by the University of Illinois Critical Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation Award in Functional Genomics to G.E. Robinson and an NSF Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.

PCR Primers

FORWARD: TAATACGACTCACTATAGG
BACKWARD: ATTAACCTCTACTAAG

Plate: BB170024A10 row: H column: 01

Seq primer: AGCGATAACAATTCACACGGA

High quality sequence stop: 619.

Location/Qualifiers

1..619

/organism="Apis mellifera"

/mol_type="mRNA"

/strain="mixed strains of European bees, predominantly A.m. ligustica"

/db_xref="taxon:7460"

/clone="BB170024A10H01"

/sex="female"

/tissue_type="brain"

/dev_stage="adult worker honey bee"

/lab_host="DH10B"

/clone_lib="Bee Brain Normalized/Subtracted Library, BB17"

/note="Organ: Brain; Vector: pT73-Pac; Site: 1: EcoRI; Site 2: NotI; This BB17 cDNA library was generated by subtraction of the BB16 library with 4000 previously sequenced clones. The BB16 library was contributed by the Soares laboratory and it was constructed and normalized as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806. RNA was prepared from dissected brains of adult worker bees of various ages and various behavioral groups."

BASE COUNT 245 a 85 c 111 g 178 t

ORIGIN

Query Match 4.6%; Score 51.6; DB 12; Length 619;
Best Local Similarity 52.8%; Pred. No. 0.059;
Matches 134; Conservative 0; Mismatches 119; Indels 1; Gaps 1;

QY 766 AAAGTCACACGAGATGTCTACCTGCCACATTTATTAGAGTGTTCACATGACATTGAA 825
|||||

DB 158 AAATCACTTTGATATGCTCATACATCAATATTATCAAGGGCTCAAGATATCTTTAA 217
|||||

QY 826 AGCATTCCTGAAGAAATCAGAAAGATAATGTTAAATTTGATATTTACTTTGTTGATCGT 885
|||||

DB 218 GCAATCTTAAGAAATAGACAGAAACATTAAGTCTTCAAAGACATGCGATTAAT 277
|||||

QY 886 GCATCTGGTCAAGAAATTCATTGGGGCAACAGATTTCTCTGCTAGTCCAGAAATTCAT 945
|||||

DB 278 GAACGTCATTATGAGGTTTGAAGGTTTAAATAAAGCTGAAAGCTGCTGCTAAATAGGT 337
|||||

QY 946 GAAGCCACAGTT-AAATTCGTTAGTTTATTAATCCACCTTAACACGGTGATATCCA 1004
|||||

DB 338 GAAGAACAGTTCAATTTGGAGAGATCTTTTGATACACCTCTCTCCACCTATGGAACCA 397
|||||

QY 1005 TCATATGAATATTA 1018
|||||

Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea; Apidae; Apis.

1 (bases 1 to 657)

Whitfield, C.W., Band, M.R., Bonaldo, M.F., Kumar, C.G., Liu, L., Pardinas, J., Robertson, H.M., Soares, B. and Robinson, G.E.

Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee

Genome Res. 12 (4), 555-566 (2002)

21929762

11932240

Contact: Gene E. Robinson
Department of Entomology
University of Illinois
505 S. Goodwin Ave., Urbana, IL 61801, USA
Tel: 217 265 0309
Fax: 217 244 3499
Email: generobi@life.uiuc.edu

This research was funded by the University of Illinois Critical Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation Award in Functional Genomics to G.E. Robinson and an NSF Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.

PCR Primers

FORWARD: TAATACGACTCACTATAGG
BACKWARD: ATTAACCTCTACTAAG

Plate: BB170018A20 row: E column: 07

Seq primer: AGCGATAACAATTCACACGGA

High quality sequence stop: 657.

Location/Qualifiers

1..657

/organism="Apis mellifera"

/mol_type="mRNA"

/strain="mixed strains of European bees, predominantly A.m. ligustica"

/db_xref="taxon:7460"

/clone="BB170018A20E07"

/sex="female"

/tissue_type="brain"

/dev_stage="adult worker honey bee"

/lab_host="DH10B"

/clone_lib="Bee Brain Normalized/Subtracted Library, BB17"

/note="Organ: Brain; Vector: pT73-Pac; Site 1: EcoRI; Site 2: NotI; This BB17 cDNA library was generated by subtraction of the BB16 library with 4000 previously sequenced clones. The BB16 library was contributed by the Soares laboratory and it was constructed and normalized as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806. RNA was prepared from dissected brains of adult worker bees of various ages and various behavioral groups."

BASE COUNT 255 a 91 c 114 g 197 t

ORIGIN

Query Match 4.6%; Score 51.6; DB 12; Length 657;
Best Local Similarity 52.8%; Pred. No. 0.06;
Matches 134; Conservative 0; Mismatches 119; Indels 1; Gaps 1;

QY 766 AAAGTCACACGAGATGTCTACCTGCCACATTTATTAGAGTGTTCACATGACATTGAA 825
|||||

Db 398 GATCAATAATATTA 411
|||||

RESULT 28

BI503620

LOCUS

DEFINITION

BI503620

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

BI503620 657 bp mRNA linear EST 08-APR-2002
BB170018A20E07.5 Bee Brain Normalized/Subtracted Library, BB17 Apis mellifera cDNA clone BB170018A20E07 5', mRNA sequence.

BI503620 1 GI:15353394

EST.

Apis mellifera (honeybee)

Apis mellifera

Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea; Apidae; Apis.

1 (bases 1 to 657)

Whitfield, C.W., Band, M.R., Bonaldo, M.F., Kumar, C.G., Liu, L., Pardinas, J., Robertson, H.M., Soares, B. and Robinson, G.E.

Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee

Genome Res. 12 (4), 555-566 (2002)

21929762

11932240

Contact: Gene E. Robinson

Department of Entomology

University of Illinois

505 S. Goodwin Ave., Urbana, IL 61801, USA

Tel: 217 265 0309

Fax: 217 244 3499

Email: generobi@life.uiuc.edu

This research was funded by the University of Illinois Critical Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation Award in Functional Genomics to G.E. Robinson and an NSF Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.

PCR Primers

FORWARD: TAATACGACTCACTATAGG

BACKWARD: ATTAACCTCTACTAAG

Plate: BB170018A20 row: E column: 07

Seq primer: AGCGATAACAATTCACACGGA

High quality sequence stop: 657.

Location/Qualifiers

1..657

/organism="Apis mellifera"

/mol_type="mRNA"

/strain="mixed strains of European bees, predominantly A.m. ligustica"

/db_xref="taxon:7460"

/clone="BB170018A20E07"

/sex="female"

/tissue_type="brain"

/dev_stage="adult worker honey bee"

/lab_host="DH10B"

/clone_lib="Bee Brain Normalized/Subtracted Library, BB17"

/note="Organ: Brain; Vector: pT73-Pac; Site 1: EcoRI; Site 2: NotI; This BB17 cDNA library was generated by subtraction of the BB16 library with 4000 previously sequenced clones. The BB16 library was contributed by the Soares laboratory and it was constructed and normalized as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806. RNA was prepared from dissected brains of adult worker bees of various ages and various behavioral groups."

BASE COUNT 255 a 91 c 114 g 197 t

ORIGIN

Query Match 4.6%; Score 51.6; DB 12; Length 657;

Best Local Similarity 52.8%; Pred. No. 0.06;

Matches 134; Conservative 0; Mismatches 119; Indels 1; Gaps 1;

QY 766 AAAGTCACACGAGATGTCTACCTGCCACATTTATTAGAGTGTTCACATGACATTGAA 825
|||||

Db 111 AAATTCCTTTGATATTGCTCATACATCAFTATTATCAAGGCTCAAGTACTCTTAA 170
Qy 826 AGCATTCCTGAAGAAATCAGAAAGATAATCTTAAATTTGATATTAATCTTTGTTATCGT 885
Db 171 GCAATTCCTTAAGAAATAGGCAAGAAACATTTACTGTTCAAAGACATGCGGATTAAT 230
Qy 886 GCATCTGCTCAGAAATTTCTATTGCTGCAACAAGATTTCTGCTAGTCCAGATATCAT 945
Db 231 GAACGTCATTATGAGGTTTGACTGTTTAAATAAAGCTGAAACTGCTGCTAATATGTT 290
Qy 946 GAAGCCACAGTT-AAATTCGTTAGTTATTAATCCACCTTAAACAGGTTGATATCCA 1004
Db 291 GAAGAACAAAGTTCAAATTTGGAGAAGATCTTTTGATACACCTCTCCACCTATGGAACCA 350
Qy 1005 TCATATGAATATTA 1018
Db 351 GATCATAAATATTA 364

RESULT 29
BI506164
LOCUS
DEFINITION BI506164 657 bp mRNA linear EST 08-APR-2002
mellifera cDNA clone BB170014B20A07 5', mRNA sequence.

ACCESSION BI506164
VERSION BI506164.1 GI:15356538
KEYWORDS EST.
SOURCE Apis mellifera (honeybee)
ORGANISM Apis mellifera
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;
Apidae; Apis.

REFERENCE 1 (bases 1 to 657)
AUTHORS Whitfield,C.W., Band,M.R., Bonaldo,M.F., Kumar,C.G., Liu,L.,
Pardinas,J., Robertson,H.M., Soares,B. and Robinson,G.E.
TITLE Annotated expressed sequence tags and cDNA microarrays for studies
of brain and behavior in the honey bee
JOURNAL Genome Res. 12 (4), 555-566 (2002)
MEDLINE 21929762
PUBMED 11932240

COMMENT Contact: Gene E. Robinson
Department of Entomology
University of Illinois
505 S. Goodwin Ave., Urbana, IL 61801, USA
Tel: 217 265 0309
Fax: 217 244 3499
Email: generobi@life.uiuc.edu
This research was funded by the University of Illinois Critical
Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation
Award in Functional Genomics to G.E. Robinson and an NSF
Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
PCR Primers
FORWARD: TAATACGACTCACTATAGG
BACKWARD: ATTAACCTCTACTAAG
Plate: BB170014B20 row: A column: 07
Seq primer: AGCGATAACAATTTCCACAGGA
High quality sequence stop: 657.

FEATURES
source
1. .657
/organism="Apis mellifera"
/mol_type="mRNA"
/strain="mixed strains of European bees, predominantly
A.m. ligustica"
/db_xref="taxon:7460"
/clone="BB170014B20A07"
/sex="female"
/tissue_type="brain"
/dev_stage="adult worker honey bee"
/lab_host="DH10B"
/clone_lib="Bee Brain Normalized/Subtracted Library, BB17"
/note="Organ: brain; Vector: pT73-Pac; Site: 1; Ecol; Site 2: Not1; This BB17 cDNA library was generated by subtraction of the BB16 library with 4000 previously

sequenced clones. The BB16 library was contributed by the Soares laboratory and it was constructed and normalized as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806. RNA was prepared from dissected brains of adult worker bees of various ages and various behavioral groups. "

BASE COUNT 253 a 91 c 115 g 197 t 1 others
ORIGIN
Query Match 4.8%; Score 51.6; DB 12; Length 657;
Best Local Similarity 52.8%; Pred. No. 0.06;
Matches 134; Conservative 0; Mismatches 119; Indels 1; Gaps 1;

Qy 766 AAAGTCACAGCAGATAGTCTACCTGCCACATATTAGAGTGTCTCAATGACCATTTGAA 825
Db 111 AAATTCATTTTGTATTTGCTCATATCATTTATTAAGGGCTCAAGTACTCTTAA 170
Qy 826 AGCATTCCTGAAGAAATCAGAAAGATAATTTTAAATTTGATATTAATCTTTGTTATCGT 885
Db 171 GCAATTCCTTAAGAAATAGGCAAGAAACATTTACTGTTCAAAGACATGCGGATTAAT 230
Qy 886 GCATCTGCTCAGAAATTTCTATTGCTGCAACAAGATTTCTGCTAGTCCAGATATCAT 945
Db 231 GAACGTCATTATGAGGTTTGACTGTTTAAATAAAGCTCAAACTGCTGCTAATATGTT 290
Qy 946 GAAGCCACAGTT-AAATTCGTTAGTTATTAATCCACACCTTAAACAGGTTGATATCCA 1004
Db 291 GAAGAACAAAGTTCAAATTTGGAGAAGATCTTTTGATACACCTCTCCACCTATGGAACCA 350
Qy 1005 TCATATGAATATTA 1018
Db 351 GATCATAAATATTA 364

RESULT 30
BI504612
LOCUS
DEFINITION BI504612 669 bp mRNA linear EST 08-APR-2002
mellifera cDNA clone BB170025B10C09 5', mRNA sequence.

ACCESSION BI504612
VERSION BI504612.1 GI:15354986
KEYWORDS EST.
SOURCE Apis mellifera (honeybee)
ORGANISM Apis mellifera
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;
Apidae; Apis.

REFERENCE 1 (bases 1 to 669)
AUTHORS Whitfield,C.W., Band,M.R., Bonaldo,M.F., Kumar,C.G., Liu,L.,
Pardinas,J., Robertson,H.M., Soares,B. and Robinson,G.E.
TITLE Annotated expressed sequence tags and cDNA microarrays for studies
of brain and behavior in the honey bee
JOURNAL Genome Res. 12 (4), 555-566 (2002)
MEDLINE 21929762
PUBMED 11932240

COMMENT Contact: Gene E. Robinson
Department of Entomology
University of Illinois
505 S. Goodwin Ave., Urbana, IL 61801, USA
Tel: 217 265 0309
Fax: 217 244 3499
Email: generobi@life.uiuc.edu
This research was funded by the University of Illinois Critical
Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation
Award in Functional Genomics to G.E. Robinson and an NSF
Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
PCR Primers
FORWARD: TAATACGACTCACTATAGG
BACKWARD: ATTAACCTCTACTAAG
Plate: BB170025B10 row: C column: 09
Seq primer: AGCGATAACAATTTCCACAGGA
High quality sequence stop: 669.
Location/Qualifiers


```
source
1. .669
/organism="Apis mellifera"
/mol_type="mRNA"
/strain="mixed strains of European bees, predominantly
A.m. linguistics"
/db_xref="taxon:7460"
/clone="BB170025B10C09"
/sex="female"
/tissue_type="brain"
/dev_stage="adult worker honey bee"
/lab_host="DH108"
/clone_lib="Bee Brain Normalized/Subtracted Library, BB17"
/notes="Organ: brain; Vector: pT73-Pac; Site 1: EcoRI;
Site 2: NotI; This BB17 cDNA library was generated by
subtraction of the BB16 library with 4000 previously
sequenced clones. The BB16 library was contributed by the
Soares laboratory and it was constructed and normalized
as described by Bonaldo, M.F., Lennon, G. and Soares,
M.B. (1996), Genome Research 6(9): 791-806. RNA was
prepared from dissected brains of adult worker bees of
various ages and various behavioral groups."
BASE COUNT 262 a 92 c 118 g 197 t
ORIGIN

Query Match 4.6%; Score 51.6; DB 12; Length 669;
Best Local Similarity 52.8%; Pred. No. 0.061;
Matches 134; Conservative 0; Mismatches 119; Indels 1; Gaps 1;

Qy 766 AAAGTCACGAGATGCTACCTGCCACATTTATAGAGTGGTTCANAGACCATTCGA 825
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 154 AAATTCACCTTTTGATATTGCTCATACATCATATTATCAAGGGCTCAAGTACTCTTAAA 213
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 826 AGCATTCCTGAAGAAATCAGAAAAGATAATGTTAAATTTGATATTACTTTGTTGTCATCGT 885
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 214 GCAATTTCTTAAGAAATAGGACAGAAAACATTACTGTTCAAAGACATGCGATTAAT 273
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 886 GCATCTGGTCAAGAAATTTCTATTGGTGCAACAAGATTTCTGCTAGTCCAGAATATCAT 945
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 274 GAAGTCATTATGCGAGGTTTGACTGGTTTAAATAAAGCTGAAGAACTGCTGCTAAATATGGT 333
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 946 GAAGCCACAGTT-AAATTCGTTAAGTTTATAAATCCACACCTTAACACGCGTGAATCCA 1004
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 334 GAAGAACAAAGTCAAAATTTGGAGAAGATCTTTTGATACACCTCTCCACCTATGGAACCA 393
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1005 TCATATGAATATTA 1018
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 394 GATCATAAATATTA 407
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: January 10, 2004, 08:39:36
Job time : 2789 secs

THIS PAGE BLANK (USPTO)